THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

THE BELL SYSTEM TECHNICAL JOURNAL

Functional Design of a Stored-Program Electronic Switching System

By HOWARD N. SECKLER and JOHN J. YOSTPILLE

(Manuscript received July 30, 1958)

The design of an experimental electronic switching system presented many new problems in specifying the functions to be performed and in developing the means to carry them out. This resulted in the use of a number of techniques new to the telephone switching field. This paper, divided into two parts, describes some of these. Part One deals with a specification of what the system must do and with its general mode of operation; Part Two describes a control philosophy employing a stored program.

PART ONE

I. INTRODUCTION

In the development of a large system, one of the most important problems is completely specifying its internal organization. The task is a complex one in the case of a telephone switching system because of the multiplicity of logical interactions involved. This paper discusses this problem as applied to an experimental electronic telephone switching system developed by Bell Telephone Laboratories.*

During the course of designing the system many techniques new to the telephone field were employed. These include a stored-program con-

---

* The system was described in a paper[1] in the previous issue which provides pertinent background material for this article.

trol philosophy and separation of the basic system functions, such as memory and logic, into efficient functionally concentrated blocks of equipment.

The experimental system, a prototype of a practical telephone office, is a real-time special-purpose machine. One of the ever-present problems was that of effecting a balance between real-time occupancy, amount of memory and other equipment, and complexity of system operation. In this balance good telephone service and over-all economy are the ultimate objectives.

Finally, a method of system operation had to be developed in order to process numerous types of telephone calls. An ideal solution is one where there is no interdependence between equipment requirements and telephone call functions. This would give rise to a universal system in which the addition or deletion of features or traffic in a telephone office would require no basic equipment changes. Within limits, this ideal solution was approached in the experimental system.

II. BASIC SYSTEM ORGANIZATION

The experimental system is composed of six major components: the switching network, the scanner, the barrier grid store, the flying spot store, the signal distributor and the central control. They are shown in Fig. 1 with their functional interconnections.

The function of the switching network[2] is to provide voice transmission connections between telephone customers; it can be considered as the primary output of the system. The switching network is made up of two main parts: the distribution switching network and the concentrators. Each customer line terminates on a concentrator, which in turn has access to the distribution switching network. Connections are established in the switching network by means of orders given sequentially to the two network controls, the concentrator marker and the distribution marker. The functions assigned to the switching network are kept as simple as possible to minimize the amount of network control equipment required. Network actions are requested by one of three orders (connect, release or trace), accompanied by appropriate terminal addresses. The network markers report the outcome of each order.

The scanner[3] is an input circuit. It is capable, on a one-at-a-time basis, of determining the state ("on-hook" or "off-hook") of each line and trunk in the office and is used to gather supervisory and dialed information. A particular customer line is interrogated by directing the scanner to a specific address; the scanner, in turn, gives an answer corresponding to the state of the line.

The barrier grid store[4] provides bulk temporary (erasable) memory for the system. It is used to store dynamic records such as line and trunk busy-idle states and dialed information. This information is stored in binary form and is read one bit at a time with random address access. A particular bit can be read and rewritten by appropriately addressing the store and giving one of four possible orders: read and regenerate, read and change, read and write zero or read and write one.

The flying spot store[5] is a bulk semipermanent memory file. It holds the telephone translation records and the operational program, and information is read out of the store on a word-organized basis. There are two general types of orders that may be given to the flying spot store: read the next program order according to a predetermined sequence or read an order at any specified location. These are known as the "advance" and "transfer" orders, respectively.

The signal distributor[6] is an output circuit that is capable of operating any one of a number of flip-flops which in turn may operate an electromechanical relay. Its prime use is for controlling relays in trunk circuits that communicate with conventional electromechanical systems. It provides compatibility between the high-speed electronic system and lower-speed mechanical systems.
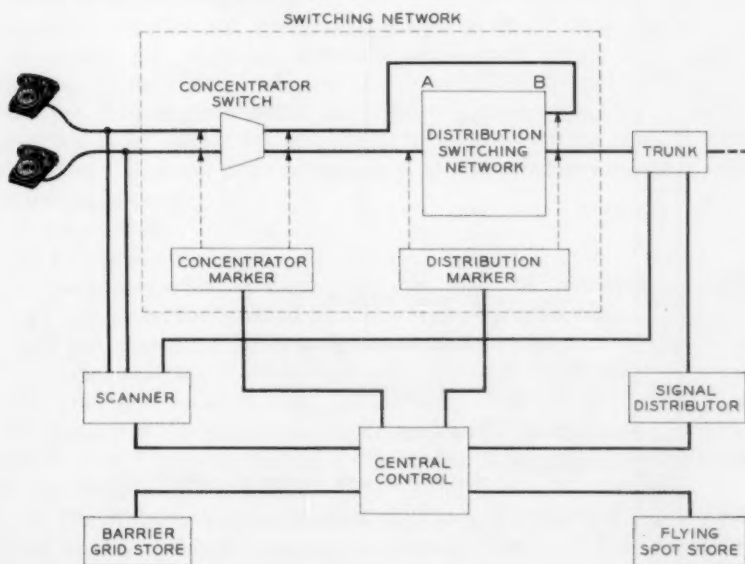
Fig. 1 — Block diagram of experimental electronic switching system.

The central control logic and organization is described in detail in Part Two of this paper. It is a logic circuit which responds to the operational program stored in the flying spot store. It controls all of the major system components, makes decisions (usually binary) based on answers it receives from them and selects the next program order to be executed.

III. SPECIFICATION OF SYSTEM SEQUENCES

Telephone switching systems are required to perform a large number of complex telephone call and administrative functions. It is necessary to specify these functions and the method of accomplishing them in exacting detail to integrate them into the system and to expedite the

RN	REQUEST NETWORK (1)
OGC	OUTGOING CALL (1)
PS PD	PERMANENT SIGNAL AND PARTIAL DIAL TIMER (1)
AIT	ABANDONED CALL AND INTERDIGITAL TIMER (1)
DT	DIAL TONE (1)
7D	7TH DIGIT STORAGE (4)
6D	6TH
5D	5TH
4D	4TH
3D	3RD
2D	2ND
1D	1ST DIGIT STORAGE (4)
DLC	DIGIT LOCATION COUNTER (3)
PC	PULSE COUNTER (4)
LL	LAST LOOK (1)
OLE	ORIGINATING LINE EQUIPMENT NUMBER (12)
A	ACTIVITY (1)

Fig. 2 — Originating register layout.

preparation of a (stored) program. Conceptually, the program itself could be written directly from broad requirements. However, it was found that a link was needed between the required external features and the specific means for implementing them in the experimental system. This link was provided in the form of detailed "sequence charts."

Before system sequence charts can be written, a language must be established. Since the charts describe the operation of a specific sy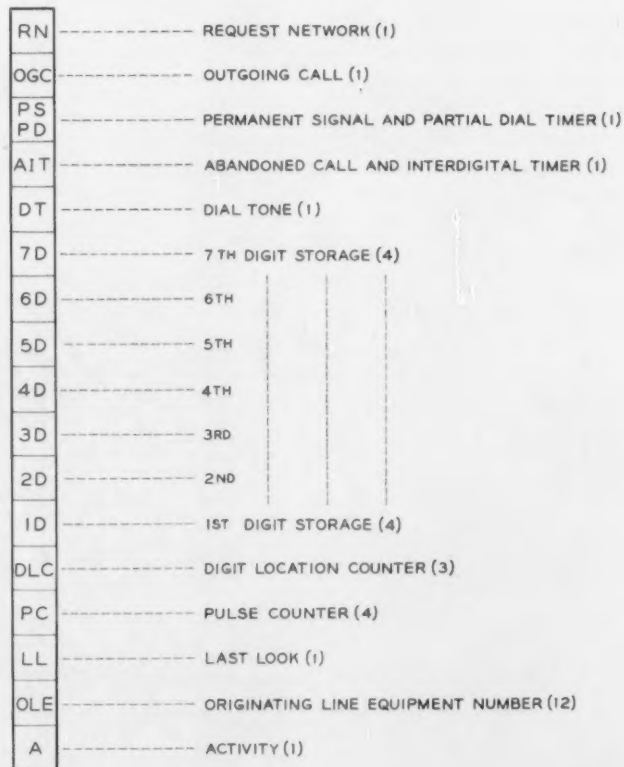stem, their form is somewhat tailored to the characteristics of that system. In the design of the experimental system sequences, the operators Read, Write, Transport and Add, abbreviated R, W, T and A respectively, were found to be sufficient to satisfy a great majority of the requirements. These operators generally act on some specified memory element or system component and, in combination with the identity of the memory element or system component, form a sequence-chart "order." A typical sequence-chart order might be: *Read the output of the scanner*, abbreviated R(S).

It is convenient to refer to locations, units of equipment and other items in abbreviated form on sequence charts. As described in the previous issue[1], barrier grid store memory is allocated in the form of "registers" performing specific functions. For example, Fig. 2 shows the layout of an originating register, which provides the temporary memory required for a customer while a call is being dialed. Storage locations in the barrier grid store are generally represented by an abbreviation designating the register followed by a hyphen and the abbreviation for the spot (bit) or spots within the register. For example, OR-LL means the "last look" spot in the originating register. The prefix may be omitted when the identity of the register is obvious. Other commonly used abbreviations are:

S for scanner,

D for signal distributor,

N as a prefix for flip-flops associated with the switching network,

FF as a prefix for specified flip-flops in the central control,

CC for unspecified flip-flops in the central control.

The four basic types of orders, described in more detail are:

1. The Read order has the general form of R (X) (Y) where X is the identity of the information to be read and Y is its storage location. Either X or Y may be omitted if not needed. A decision is always made as the result of a read order.

2. The Write order has the form $W(K) \rightarrow (Z)$, where K is a specific fixed number, either in numerical or symbolic form, and Z is the storage location into which the information K is to be written.

3. The Transport order has the general form $T(X)$ $(Y) \rightarrow (Z)$, where X is the identity of the information at storage location Y and Z is the storage destination of X. As above, either X or Y can be omitted if not needed. This order is used to move any information from one location to another.

4. The Add order, which is also used for subtraction, is of the form $A(K) \rightarrow Z$ or $A(-K) \rightarrow Z$, where K is a positive integer and Z a storage location.

IV. SYSTEM SEQUENCES AND TECHNIQUES

Although the experimental system is limited in scope, a large number of sequence charts are needed to specify its functional operation. A few sequences followed at the start of a telephone call are discussed in this section in order to illustrate how certain typical functions are accomplished and what techniques are used. For the sake of brevity, many details have been omitted, particularly in the latter part of this section.

4.1 *Supervisory Functions*

One of the functions of a telephone switching system is the recognition of supervisory signals from customers, that is, new requests for service and terminations of established calls. The customer signals the telephone office by lifting his telephone receiver from the switchhook (off-hook) to request service and later, by placing it back on the switchhook (on-hook), to terminate service. These actions respectively close and open a metallic path to the telephone office. Table I shows the scanner output for each state.

Since the scanner can only determine the existing state of a customer line, some memory must be provided in order to detect changes in line state. This is done by assigning two barrier grid store bits to each line. These are referred to here as the first and second line memory bits, or L1 and L2 spots. The experimental system uses three of the four possible combinations of each pair of line memory bits, with assignments as shown in Table II. The "served by a register" state is assigned when

TABLE I — LINE STATES AND CORRESPONDING SCANNER OUTPUTS

Telephone Receiver	Telephone Line	Scanner Output
on-hook	open	0
off-hook	closed	1

TABLE II — LINE MEMORY ASSIGNMENT

| | Barrier Grid Store Memory Bits | |
Status of Line	L1	L2
Idle.	0	0
Talking.	1	0
Served by a Register.	0	1

special action is taken on a line such as dial pulse recording, ringing or disconnect timing.

The experimental system starts action on supervisory signals from customers within 100 milliseconds of their occurrence. Fig. 3 illustrates, in sequence chart form, the work that may be done on each customer line once every 100 milliseconds; reference to Table III will be helpful in understanding the description that follows.
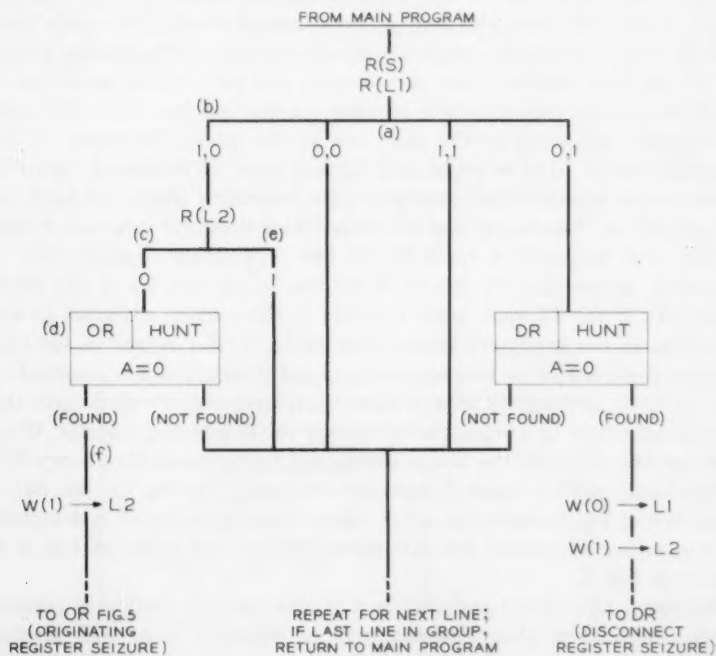

Fig. 3 — 100-millisecond supervisory line scan.

TABLE III — LINE SIGNALS RECOGNIZED IN SUPERVISORY SCAN
SEQUENCE BY COMPARING SCANNER OUTPUT WITH LINE SPOTS

Scanner Output	Line Spots		Customer Signal
	L1	L2	
0	0	0	None
1	1	0	None
1	0	0	Call Originated
0	1	0	Disconnect
0 or 1	0	1	Line under control of other sequences

As described in a previous paper[1] the "main program" is the vehicle
which assigns the multiplicity of real-time functions which the system
must perform at various periodic intervals. At the start of the super-
visory scan for originations and disconnects, the scanner is directed to
the first line of a predetermined group and the barrier grid store is ad-
dressed to read its associated L1 spot; this is specified by the symbols
R(S) and R(L1) near the top of the sequence chart. This leads to a
branch (point a) in the sequence which opens to four possible paths.
If the answers received from the scanner and barrier grid store match
(0,0 or 1,1), no further action is taken on the line (see Table III) and
the system advances to the next line in the group. However, if the
combination (1,0) is returned, the L2 spot must be examined (point b)
to determine whether it is in the zero state, indicating that a call has been
originated by this line in the previous 100-millisecond interval. If this
is the case (point c), a hunt for an idle originating register (OR) is
initiated, as specified by the symbolic box in the left leg of the chart
(point d). If the L2 spot reads 1 (point e), the system advances to the
next line in the group, as above. Proceeding to the output of the idle
register hunt box, if an idle register is found (point f), a 1 is recorded in
the L2 spot, putting the line in the served-by-a-register state, and the
system advances to a sequence for seizing the originating register. If no
idle register is found, the above action will be repeated once every 100
milliseconds until a register becomes available. The leg on the right-
hand side of Fig. 3 shows the action taken when a disconnect is detected.
The sequence of system actions represented by the boxes of Fig. 3 is
shown in Fig. 4.

Although the above example is a simple one, it illustrates certain
fundamental principles. The majority of customer lines will always
be in the quiescent states of idle (0,0) or talking (1,0), and only a small
amount of system action is required for them once every 100 milli-

seconds. Only when a line changes state does the system have a complex job to perform. Since the number of changes of line state per 100-millisecond interval is very small, the percentage of real time consumed by the 100-millisecond line-scan sequence is correspondingly small. Another factor to be considered is the amount of memory required for each function. Here two barrier grid store (BGS) bits are assigned to each line and a certain amount of program space in the flying spot store is used for the supervisory scan function. It is apparent after some study that a decrease in the amount of BGS memory assigned to each line will result in an increase in real-time consumption and an increase in program space. To show that a good balance, in this respect, has been attained in the experimental system is beyond the scope of this article.

4.2 *Originating Register Seizure*

In the experimental system, BGS memory is associated with certain telephone functions in a manner somewhat similar to the use of common control equipment such as registers, senders and markers in conventional relay telephone switching systems. As mentioned previously, the originating register illustrated in Fig. 2 shows the BGS memory that is associated
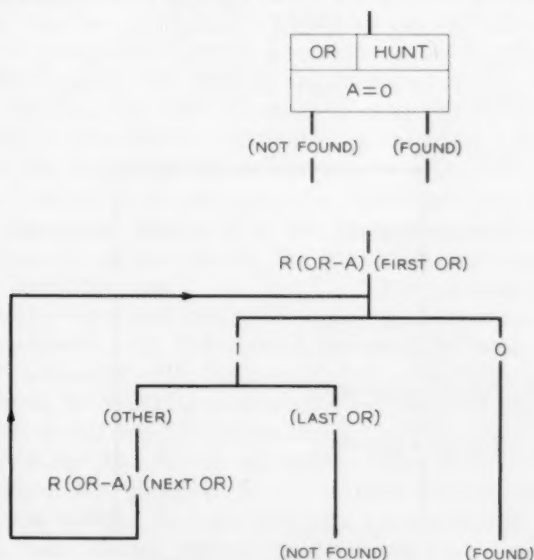

Fig. 4 — Register hunt.

with a line during dialing. For each bit or subgroup of bits, the functional abbreviation, the full name and the number of bits (in parenthesis) are given. The originating register is one type of a class of BGS registers referred to as "call registers." In this system a more than sufficient number of call registers to give a good grade of service can be economically provided, since no equipment other than the memory itself is affected by the number of registers.

After the origination of a new call has been detected and an idle originating register is found, as shown in Figs. 3 and 4, the originating register must be seized and action started to connect a dial tone trunk to the line via the concentrator and distribution network. The system sequences for carrying out this task are shown in Fig. 5. While a register is idle the only useful information recorded in it is a 0 in the "activity" (busy-idle) spot. This eliminates the need for keeping the remaining register memory bits in a fixed state over long periods of inactivity, as
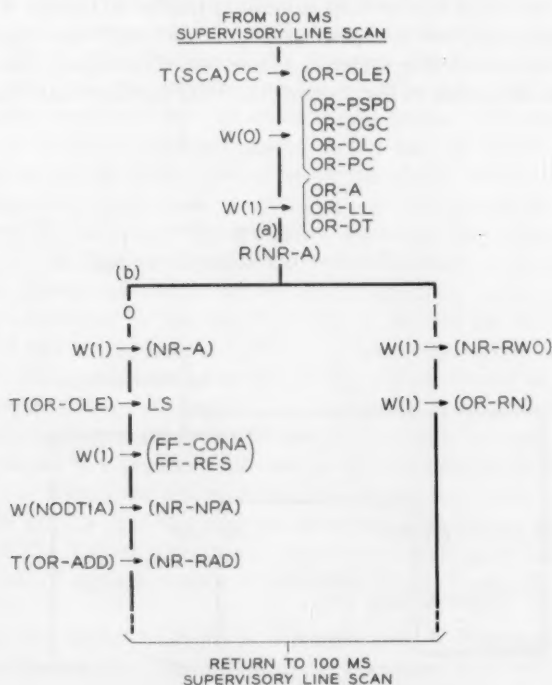

Fig. 5 — Originating register seizure.

would be the case for the last-choice registers in a large group. Therefore, when a register is first seized various state and reference information must be recorded, as shown by the Write orders in the top center leg of Fig. 5.

4.3 *Dial Tone Connection*

The switching network operates at a speed of approximately one millisecond per task, while the rest of the system performs operations at a microsecond rate. For this and other reasons buffer memory associated with the switching network is provided in the BGS and has been designated as the "network register" (NR). The network register is one type of a second class of BGS registers which are referred to as "control registers." Usually only a single control register is needed for a particular function such as network control.

After the originating register is seized, the network register activity spot (NR-A) is read as shown in Fig. 5. If a 1 is read, indicating that the network is busy, a request for later action is recorded in both the originating and network registers. If a 0 is read, indicating that the network is available, the network register is seized. In this case, the line equipment number (OR-OLE) is passed to the network line selector (LS) in the concentrator marker and certain control flip-flops are set to start connecting the line through the concentrator to a distribution network terminal. The type of action required (NODT1A—first part of dial tone connection) and address of the originating register (OR-ADD) are also recorded in the network register so the system will know where to continue in the program when this segment of the network action is completed. Having done this, the system returns to sample the condition of the next line in the 100-millisecond supervisory line scan.

Periodically the system checks the switching network to determine if it has completed a task. This is done by reading a "break in" flip-flop (FF-NBI) associated with the network control circuits, as shown in Fig. 6. When the switching network has completed an assignment, its own control circuit sets FF-NBI so that a 1 will be read the next time it is interrogated. This in turn will cause a break in the main program and will direct the system to determine, from information previously stored in the network register, what task the network has completed and then, from answers received from network control circuits, what action should be taken next. The remainder of the dial tone connection is established in this way.

The "simple-task-at-a-time" method of operating the network permits efficient utilization of the common system memory and central control, thus minimizing the quantity of equipment individual to the network. It also provides a means of coupling the lower-speed network to the rest of the system.

4.4 Detection of Dial Pulses

After a line has been assigned to an originating register and a dial tone connection is established, means must be provided for detecting and storing dial pulses. This is accomplished by the sequence shown in Fig. 7. The main program directs the system to carry out this sequence for all active originating registers once every 10 milliseconds, as specified by the symbolic box shown at the top of the figure. The pulses generated by the customer's dial are a series of momentary "on-hook" intervals, the number corresponding to the digit dialed, except for zero which is represented by 10 pulses. The 10-millisecond scanning rate is fast enough to insure that every legitimate dial pulse and the "off hook" intervals that separate them will be sampled at least once[3].

After the originating line equipment number (OLE) recorded upon seizure of the originating register has been read as indicated in Fig. 7, the scanner is directed to scan the line and the BGS is addressed to read the last look spot (LL). This spot contains an up-to-date record of the state of the line according to the last previous scanner reading (0 for "on-hook" and 1 for "off-hook"). If the answers received from the scanner and BGS match (0,0 or 1,1) no change has taken place since

FROM MAIN PROGRAM

R(FF-NBI)

0   1

TO MAIN PROGRAM

TRANSFER TO
PROGRAM ADDRESS
IN NR-NPA
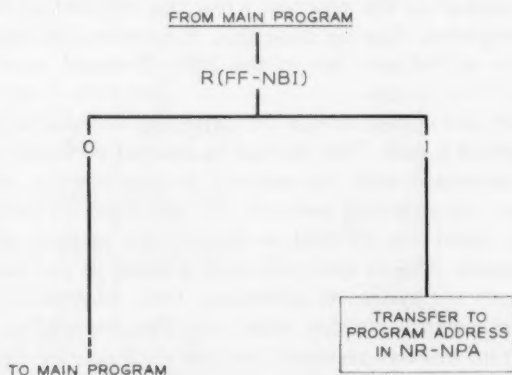
Fig. 6 — Periodic network check.

the last scan and the program proceeds to the next originating register. However, if the scanner reads 0 and the LL spot 1, this is interpreted as the leading edge of a dial pulse; in this case, the LL spot is put up to date (0) and the pulse counter spots (PC) are incremented by one. In addition, the abandoned call and interdigital timer spot (AIT) is written to 0, thus permitting another sequence (not shown) to detect an abandoned call if the on-hook condition persists. The dial tone spot (DT) is also read to determine if dial tone is connected to the customer line. Dial tone is disconnected from the line as soon as the first pulse of the first dialed digit is received, by a method similar to the one

Fig. 7 — Originating register dial pulse scan.

Fig. 8 — General interrelationship of system sequences.

1340

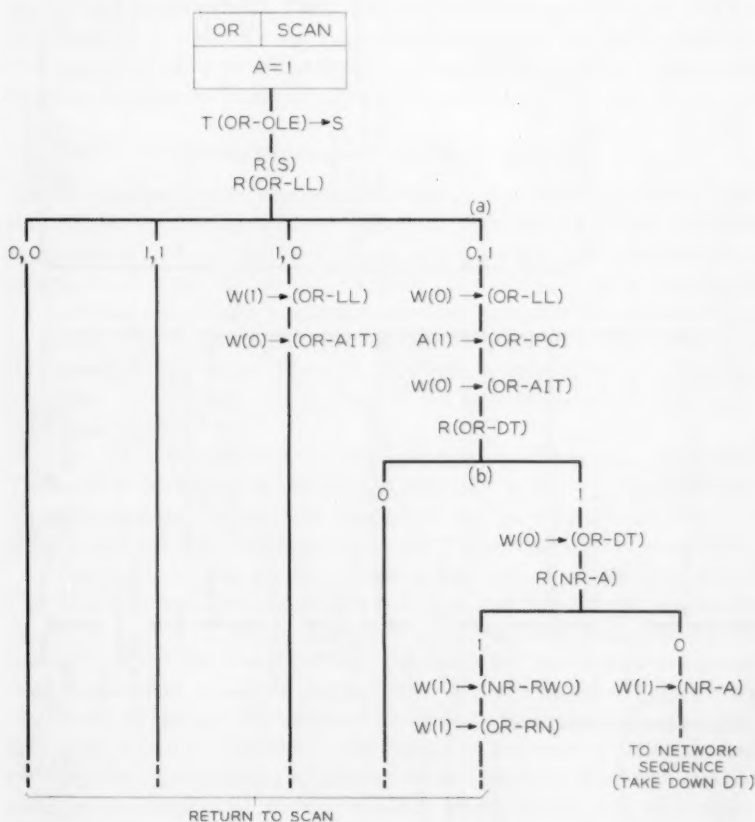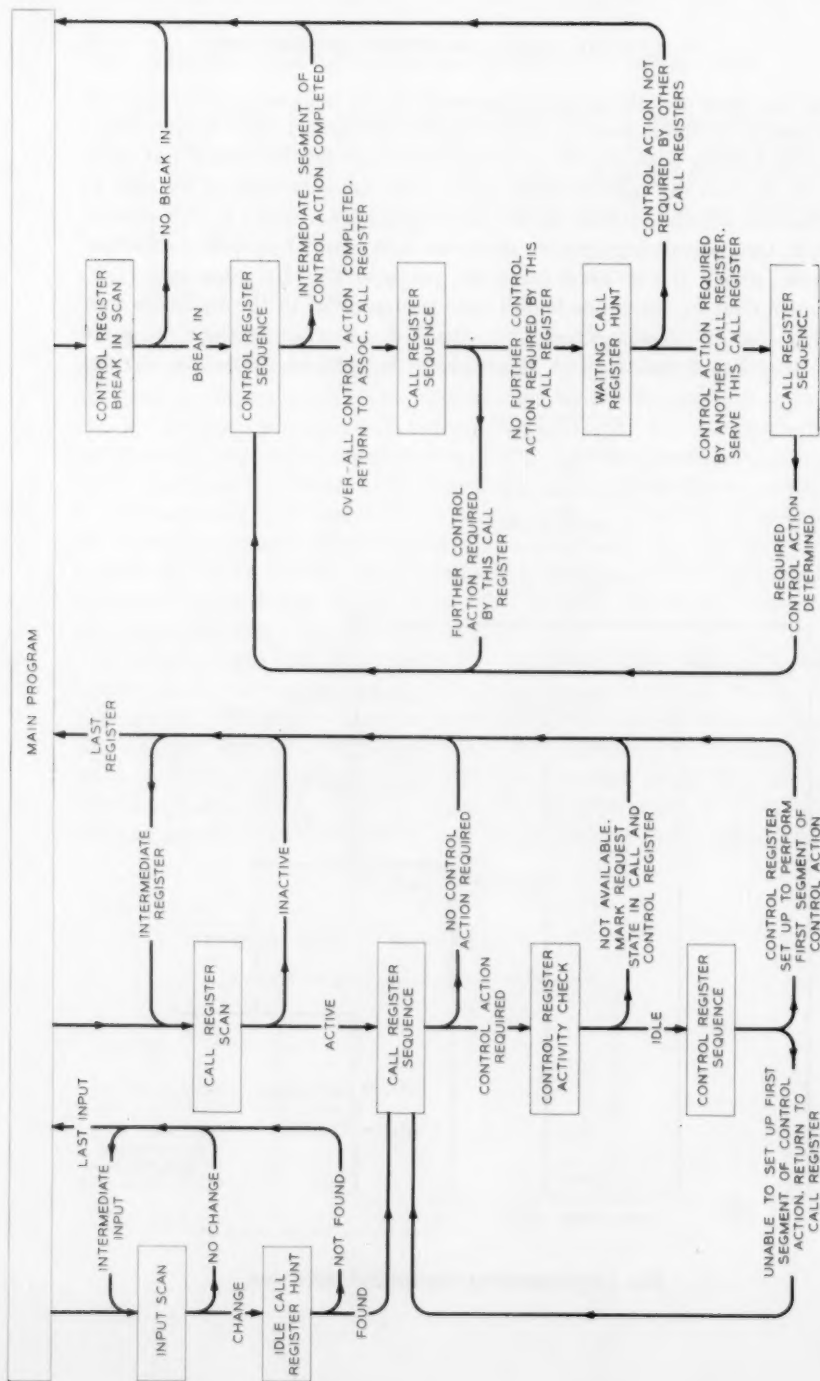previously discussed for making a network connection. The remaining leg of Fig. 7, when the responses from scanner and BGS are (1,0) respectively, is followed after the trailing edge of a dial pulse is detected.

This sequence (Fig. 7) only accomplishes the detection and storage of dial pulses. Another one recognizes the end of each digit, stores the accumulated pulse count in the appropriate digit storage slot and prepares the originating register for the next digit. Still other sequences carry out the various tasks required to complete the call.

The previous examples are typical of the multiplicity of call sequences or flow charts that are necessary to describe and specify the system operational requirements. They provide a convenient tool which permits the designer to optimize the use of equipment and real time, keep an intelligible record of what the system does and also put the requirements into a form that can be readily programmed.

V. GENERAL INTERRELATIONSHIP OF SYSTEM SEQUENCES

From the few system sequences illustrated thus far, it is evident that the experimental system is not equipped with specific control circuits or programs that are individually associated with a telephone call as it progresses through the system. Instead, the necessary control functions are provided by a single common equipment unit on a time-shared basis as directed by a hierarchy of programs performing a multiplicity of fundamental operations. There is, however, a general mode of system operation which clearly emerges out of the interrelationships among the system sequences.

There are several classes of jobs specified by the system sequences. The more important ones are listed in Table IV with a brief explanation. The relationships among these tasks form the basic operational pattern of the experimental switching system. Fig. 8 illustrates this in block form.

A detailed example of the system action taken, if the left leg of Fig. 8 is followed, can be obtained by tracing through the sequences shown in Figs. 3, 4 and 5. The main program periodically directs the system to determine if any calls have originated or terminated in a group of customer lines, using an "input scan" (top of Fig. 3). If a particular line shows no change, the sequence proceeds to interrogate the next line, and so on. When the last line in the group has been served, the sequence reenters the main program to determine the next task required. Should a change be found on a line, an "idle call-register hunt" (Fig. 4) is made. If no idle register is found the sequence proceeds to the next line or the main program. If, as in the usual case, an idle call register is found, a "call register sequence" is entered (up to point a on Fig. 5). At this point,

TABLE IV — FUNCTIONAL CLASSIFICATION OF SYSTEM SEQUENCES

Sequence	Description	Example
Input Scan	A search or scan of system inputs for new work	Top of Fig. 3 — $R(S)$ $R(L1)$
Call Register Scan	A search or scan of call registers (e.g. originating register) for work	Box at top of Fig. 7
Call Register Sequence	A specialized task associated with a call register	Fig. 7
Control Register Break-In Scan	A periodic examination of an equipment for control action	Fig. 6
Control Register Sequence	A task associated with a control register (e.g. network register) and its corresponding equipment unit	Bottom half of Fig. 5
Main Program	The sequence which provides priorities and timing in assigning all system tasks	—
Idle Call Register Hunt	A search for an idle call register	Fig. 4

control action, such as network orders, may be required. The system then proceeds to check the control register for availability and, if appropriate, it advances to a control register sequence (starting at point b, Fig. 5). On some sequences a decision will determine whether control action is required (see point b on Fig. 7); on others, it is required unconditionally (point a, Fig. 5). The description of the remainder of Fig. 8 is left to inspection. This method of system operation was an outgrowth of efforts to optimize the solution of a specialized problem, but many of the techniques employed can be, and some have been, employed in other real-time machines.

The material presented thus far has illustrated what the system is to do, and how it should be done in terms of the external characteristics of the system equipment units. The specific means for implementing the actions prescribed by the system sequence charts, that is, the control philosophy, is the subject of Part Two of this article.

PART TWO

VI. INTRODUCTION

In developing a control philosophy to govern the sequences of system actions, at least two major alternatives must be considered. The first is the use of electronic logic circuits to cause the required actions to be carried out in proper sequence, similar to the way in which relay circuits control present switching systems. This would involve, to a large extent, the individual tailoring of these logic circuits to requirements specified by

the sequence charts described in Part One. The second alternative is realization through a stored program, whereby the pattern of sequences would be incorporated into the program, with the control circuitry designed to interpret and carry out the different types of basic program orders.

In the digital computer field stored programs are used where flexibility is of paramount importance, because of the wide range of applications general-purpose computers must have. Since a telephone switching system is by nature a special-purpose rather than a general-purpose system, the choice between the circuit logic and stored-program approaches was not immediately obvious. The choice of the latter was initially based largely on expected economies of control circuitry and removal of the dependence of the circuit logic on the required sequences of actions. It became increasingly clear, as time progressed, that a still more important aspect of stored-program operation is the flexibility it affords. Of immediate significance in this respect is the ease of addition and revision of system functions; of long-range significance is the real possibility of a universal control for telephone switching systems which could adapt them for local, tandem or toll use as dictated by the stored program.

A stored program for an electronic telephone switching system is a set of encoded orders specifying exactly what the system must do at all times under all possible customer input situations. It exercises exclusive control over the entire system. The detailed functional design of a stored-program electronic switching system, after the control philosophy and required sequences have been established, may be divided into three phases: (1) enumerating the types of basic operations to be incorporated as program orders and determining an encoding structure for these, (2) designing the logic circuitry for their interpretation and execution and (3) writing programs containing these orders to fulfill the requirements of system operation as set forth on the sequence charts. These phases are developed in the following sections.

VII. DEVELOPING A PROGRAM ORDER STRUCTURE

7.1 *Enumeration of Required Operations*

From examination of the general pattern of system operation and the sequence charts already described, it is evident that the system must be equipped to perform certain specific operations. These may be first classified as *decision* and *nondecision* operations.

A decision operation occurs at a branch point of a sequence chart and

provides for a binary choice of alternative actions to follow it. In this category may be included:

1. Read BGS at specified address.
2. Read specified flip-flop.
3. Read scanner at specified address.
4. Match two binary words (determine whether they are equal).

A decision operation is further specified by the addition of a modifier which determines which of two alternatives is to correspond to each of the two possible results. One of the alternative actions following the decision point will be to continue with the next operation appearing sequentially in the program. The other will be to transfer control to some other (predetermined) location in the program.

A nondecision operation performs work called for as a result of some previous decision operation. A nondecision operation is always followed directly by the succeeding operation specified in the program. Included in this category may be:

5. Write 1 or 0 in BGS at specified address.
6. Write 1 or 0 (set or reset) specified flip-flop.
7. Read BGS at specified address (nondecision) and write result in specified flip-flop.
8. Read specified flip-flop (nondecision) and write result in BGS at specified address.
9. Store a specified constant (usually a binary word representing an address) in a specified group of flip-flops.
10. Gate the information in one specified flip-flop group to a second specified flip-flop group.
11. Transfer control from the present program location to a specified program location.
12. Add 1 to the binary number stored in a flip-flop group.
13. Regenerate the BGS at a specified address.

7.2 *Encoding the Program Orders*

From the 13 basic types of operations listed above may be formed an encoding structure for the specific program orders yet to be derived. The process of encoding consists of determining the number of bits a program order of fixed size will contain and what the function of each bit will be.

It is apparent from the operation types that a program order generally contains an instruction (*what*) and an address (*where*). A first step in deriving a coding structure is to classify the addresses required by each operation type, as follows:

1. Read BGS. The BGS contains a square array of 128-by-128 stor-

age locations. Access to one of these therefore requires an address of 14 bits. This erasable bulk memory is organized so that call registers and other aggregates of temporarily stored information are arranged generally within either columns or rows of the array. In this system the BGS will thus be addressed most frequently on successive usages to locations within a single column or row. Therefore, it is logical to precede such a series of BGS operations by presetting a constant row or column address in flip-flop group memory. This will require program orders dealing with the BGS to contain only a seven-bit column or row address rather than a full 14-bit address. Most BGS reading orders will thus contain a seven-bit address, the other seven bits being stored in flip-flop memory.

2. Read Flip-Flop. The size of address required in this type of operation depends only on the anticipated number of flip-flops which must be individually read. Since this number will be in the range between 64 and 128, a seven-bit address is required.

3. Read Scanner. It is required that an order to read the scanner output be preceded by storing the desired address in a flip-flop register associated with the scanner. The Read Scanner operation itself contains no address.

4. Match. This operation determines whether two binary words stored in flip-flop groups are equal. A simple approach to this is to require that one of the two words be stored in a reference flip-flop group with which a matching circuit is associated. The other word may be stored in any one of a number of other flip-flop groups. The matching operation then must include only the address of the latter flip-flop group. Because it was determined that between 16 and 32 flip-flop groups are required for various parallel-access, short-term storage functions, a Match operation requires a five-bit address.

5. Write 1 or 0 in BGS. Same as Read BGS (seven-bit address).

6. Write 1 or 0 in Flip-Flop. Same as Read Flip-Flop (seven-bit address).

7. Read from BGS to Flip-Flop. This operation requires two addresses, a seven-bit partial BGS address (the other seven bits preset in a flip-flop group) and a flip-flop address. For circuit economy, however, 32 of the maximum of 128 individually addressable flip-flops (item 2 above) will suffice to store directly information taken from the BGS. Thus a five-bit flip-flop address accompanies the seven-bit BGS address.

8. Read from Flip-Flop to BGS. Same as above (seven-bit BGS address and five-bit flip-flop address).

9. Store Constant in Flip-Flop Group. It is convenient in this case to

consider the constant as the address-part of this operation. For a number of reasons, involving mostly the detailed numbering plan of the entire system, most of the flip-flop groups contain 14 flip-flops. This operation, if accomplished by a single program order, would therefore contain a 14-bit "address". Since the operation will be restricted to only a small number of flip-flop groups for circuit economy, it is convenient to consider the identity of the flip-flop group in which the constant is to be stored as part of the instruction rather than as an address.

10. Gate Flip-Flop Group to Flip-Flop Group. This operation contains two five-bit flip-flop group addresses.

11. Transfer. An operation causing an unconditional transfer to some point in the program may (a) contain the full flying spot store address to which the transfer is to be made, or (b) specify the identity of a flip-flop group which contains the transfer address. In the latter case, a five-bit address is required; in the former, a 14-bit address is required, assuming the program to be limited to a portion of the flying spot store plates so that 14 bits will afford complete access.

12. Add 1. If we assume that, for circuit economy, a parallel Add 1 circuit is associated with one particular flip-flop group rather than with all flip-flop groups, then the Add 1 operation will cause the binary word contained in that flip-flop group to be incremented by one. Thus, no address is contained in this operation.

13. Regenerate BGS. Since the BGS will be periodically regenerated by rows or columns, this operation need contain only a seven-bit address, the other seven bits being previously stored in a flip-flop group.

TABLE V

Operation Type		Address Requirements	
No.	Name	Number of Addresses	Size (Bits)
1	Read BGS	1	7
2	Read Flip-Flop	1	7
3	Read Scanner	0	
4	Match	1	5
5	Write in BGS	1	7
6	Write in Flip-Flop	1	7
7	Read BGS to Flip-Flop	2	7, 5
8	Read Flip-Flop to BGS	2	5, 7
9	Store Constant	1	14
10	Gate	2	5, 5
11(a)	Transfer	1	14
11(b)	Transfer	1	5
12	Add 1	0	
13	Regenerate BGS	1	7

The address requirements just developed are summarized in Table V. With this information it is possible to establish tentatively the number of bits required by program orders to be derived from the 13 operation types. There are two operation types requiring 14-bit addresses, Transfer and Store Constant. A single program order will be derived from the Transfer operation which contains a 14-bit flying spot store address. In the case of the Store Constant operation, however, it will be expedient to permit a 14-bit constant to be stored in any one of *three* different flip-flop groups; three program orders will therefore be derived from this operation type.

There are, then, four program orders requiring a 14-bit address; thus, two additional bits must be added to distinguish among them. If 16 bits are consumed by only these four program orders, 17 is certainly a lower bound on the number of bits required in a fixed-size program word, since the remaining program orders must also be encoded. As will be seen, 17 bits is sufficient as well as necessary and is adopted as the program word size.

Proceeding with the assumption of a 17-bit word size, the information in Table V is converted to the form of Table VIA to begin the encoding process. Here the addresses included in each operation type are assigned bit positions, starting from the right-hand side. For those cases where two addresses exist (7, 8, 10), the two are placed adjacent, except for operation type 10, where the two five-bit addresses are placed as shown so that the left-hand one is lined up with those of 7 and 8 to achieve as unified a layout as possible.

TABLE VIA

	Operation Type	Program Word Bit Position																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Read BGS											≡	≡	≡	≡	≡	≡	≡
2	Read Flip-Flop											≡	≡	≡	≡	≡	≡	≡
3	Read Scanner																	
4	Match													≡	≡	≡	≡	≡
5	Write in BGS											≡	≡	≡	≡	≡	≡	≡
6	Write in Flip-Flop											≡	≡	≡	≡	≡	≡	≡
7	Read BGS to Flip-Flop						≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
8	Read Flip-Flop to BGS						≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
9	Store Constant				≡	≡						≡	≡	≡	≡	≡	≡	≡
10	Gate						≡	≡	≡	≡	≡			≡	≡	≡	≡	≡
11(a)	Transfer				≡	≡	≡	≡	≡	≡	≡	≡						
11(b)	Transfer													≡	≡	≡	≡	≡
12	Add 1																	
13	Regenerate BGS											≡	≡	≡	≡	≡	≡	≡

TABLE VIb

Operation Type		A			B		C					D						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Read BGS											▨	▨	▨	▨	▨	▨	▨
2	Read Flip-Flop											▨	▨	▨	▨	▨	▨	▨
3	Read Scanner																	
4	Match												▨	▨	▨	▨	▨	▨
5	Write in BGS											▨	▨	▨	▨	▨	▨	▨
6	Write in Flip-Flop											▨	▨	▨	▨	▨	▨	▨
7	Read BGS to Flip-Flop						▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
8	Read Flip-Flop to BGS						▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
9	Store Constant				▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
10	Gate						▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
11(a)	Transfer				▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
11(b)	Transfer												▨	▨	▨	▨	▨	▨
12	Add 1																	
13	Regenerate BGS											▨	▨	▨	▨	▨	▨	▨

Table VIb repeats the information in VIa, with the lines of demarcation bounding the addresses darkened so that the 17-bit word is subdivided into groups of three bits (called A), two bits (B), five bits (C) and seven bits (D). In places where a five-bit address occupies the rightmost positions [4, 10, 11(b)], no subdivision is shown because it will prove convenient circuit-wise to treat these as seven-bit addresses with the two most significant bits zero.

A program order word may now be thought of as containing an A code, a B code, a C code and a D code. For orders derived from operation 1 above, for example, the BGS seven-bit partial address will be contained in the D code, while the A, B and C codes together will specify the exact nature of the operation to be performed. For operation 11(a), on the other hand, the B, C and D codes together will contain the flying spot store address to which a transfer is to be made, while the A code will direct that a transfer be made.

A process of encoding has now been described except for the detailed specification of all the program orders to be derived from the 13 operation types and the illustration of how these fit into the proposed structure. This final phase will be deferred to a later section which deals with the writing of programs. In that section a complete set of program orders will be presented.

At this point we turn to a discussion of the circuit logic of the central control, whose function is to interpret and carry out all the types of 17-bit program orders provided.

VIII. THE CENTRAL CONTROL

8.1 *Basic Considerations*

The central control is the heart of the electronic switching system. Its actions are required for all system functions to be accomplished. It consists almost entirely of direct-coupled semiconductor circuitry[7] arranged in logic configurations to accomplish the interpretation and execution of each program order originating from the flying spot store (FSS).

Central control processes these orders one at a time. The time allotted to carry out an order is dependent upon the speeds of the major system components, in particular the flying spot store, barrier grid store and central control itself. The progress through the program, order by order, is controlled by a clock whose period is constant and sufficiently long to cover the operating time of any of the above-mentioned units. This time will be referred to as the "cycle time" and is of several microseconds duration.

When a program order is passed from the FSS to central control on the occurrence of a clock pulse, the central control logic circuits, within the ensuing cycle time, decode the order and prime those parts of the central control which are to participate in the execution of the order. The succeeding clock pulse causes the execution of the order and, at the same time, causes the next order in the program to pass from the FSS to the central control for similar processing, as shown in Fig. 9. This "overlap" arrangement between the FSS and central control is used to conserve system time. The order-by-order progression just described is altered in the case of decision orders, as will be discussed later.

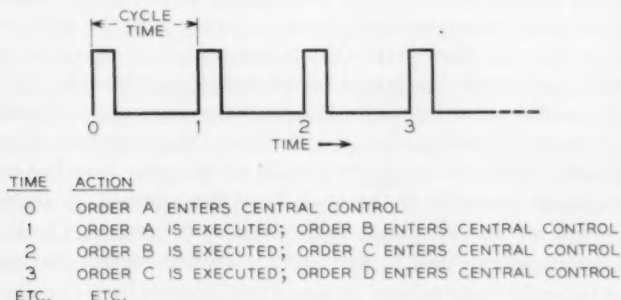| TIME	ACTION
0	ORDER A ENTERS CENTRAL CONTROL
1	ORDER A IS EXECUTED; ORDER B ENTERS CENTRAL CONTROL
2	ORDER B IS EXECUTED; ORDER C ENTERS CENTRAL CONTROL
3	ORDER C IS EXECUTED; ORDER D ENTERS CENTRAL CONTROL
ETC.	ETC.

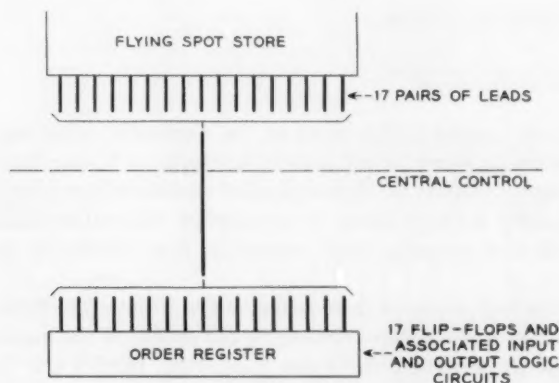Fig. 9 — Electronic switching system clock timing.

Fig. 10 — Central control order register.

Since the FSS output is detected by means of a short sampling pulse, the first requirement in central control is for a register to receive and then hold the order for a cycle time. This register is called the order register and is shown in Fig. 10. It consists of 17 flip-flops, one to hold each bit of the order, and logic circuits to properly associate the register with the circuits connected to it. The first component of the cycle time consumed during the processing of an order is the interval from the beginning of a clock pulse to the time at which the outputs of the flip-flops and amplifiers of the order register correctly indicate the bits of the order; this is approximately one-third of the cycle time.

Since the order is binary-encoded, a decoding or translating process is next required, to convert the order into a form which can be used for establishing the circuit conditions required for its execution. The association of an order translator used for this purpose with the order register is shown in Fig. 11. The input to the translator is a 17-bit order; its output usually consists of signals on a few of more than 100 leads. The translator logic circuits and amplifiers themselves require about one-third of a cycle time to respond, so that, by the time the proper translator outputs are active, approximately two-thirds of the cycle time has expired. The remaining one-third of the cycle time is consumed by those logic circuits that act in response to order translator outputs. Clock pulses control circuit actions only at the input to central control (between the FSS and the order register) and at the output stage of logic circuits where the execution of orders is accomplished.

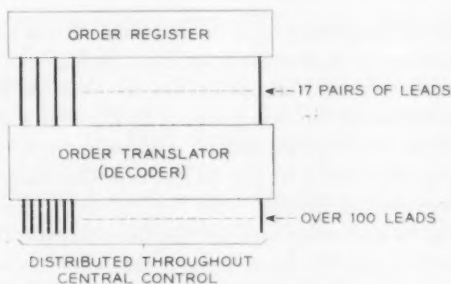The order translator outputs are connected in some manner to almost

Fig. 11 — Central control order translater.

every circuit of the central control.* Much of this circuitry consists of flip-flop registers, including special-purpose registers associated with specific functions and general-purpose registers with multifunctional application. The registers are employed mainly for short-term high-speed storage of addresses associated with the various major system units. The communication among these flip-flop groups, and between them and circuits external to central control, may be achieved through the use of a common bus, since the program orders permit only one parallel aggregate of bits to be communicated in a single cycle time. In addition to establishing conditions for communications via the bus, the order translator outputs also control actions internal to a register circuit or actions associated with circuits having control rather than register functions.

Fig. 12 shows a simplified picture of the relationships among the order register, order translator, flip-flop group registers, bus, some control circuits yet to be described and circuits external to central control. All of the registers attached to the bus need only contain 14 or fewer flip-flops if it is assumed that this number is sufficient to specify an address in any major system unit. As shown, some registers have communications both to and from the bus; all of the general-purpose and some special purpose registers are in this category. Other special-purpose registers have communications either to the bus or from the bus but not both, most of these being directly associated with external system units.

The bus itself consists of 14 pairs of OR gates whose outputs serve as the common communication path. For a 14-bit register having communication to and from the bus, the paired outputs of its 14 flip-flops can be gated to the inputs of the 14 flip-flops of any other register so connected.

---

* It is beyond the scope of this article to cover completely the many detailed functions of central control. It is intended, rather, to develop and explain only the more fundamental features.

To clarify this arrangement and to illustrate how a simple program order is carried out by central control, an example is given which involves only those circuits already discussed. Let us examine how the order "Gate the information in flip-flop group A to flip-flop group B" (operation type 10, Table V) is accomplished. The logic circuit actions resulting from this order are shown in Fig. 13. At time 0 in this example a program order enters central control from the FSS coincident with a clock pulse and is held by the order register. As time elapses, the flip-flops of the order register respond to the order and the order translator logic responds to the outputs of these flip-flops. For the order "Gate FFG A (flip-flop group A) to FFG B", two translator outputs are activated, one causing the outputs of the flip-flops of FFG A to be gated to the bus input, the other readying the bus output to be gated to the inputs of the flip-flops of FFG B. Thus, within a cycle time following time 0—that is, before the occurrence of the next clock pulse at time 1—the information held by FFG A will have been transmitted via the bus to gates at the input of FFG B. Then, at time 1, coincident with the next clock
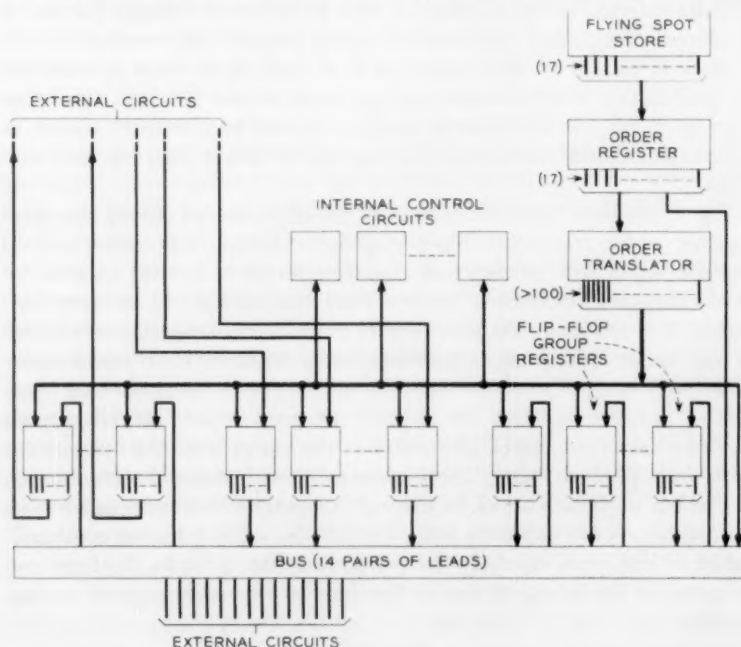

Fig. 12 — Simplified block diagram of central control.

pulse, this information is gated into FFG B and the next order enters the order register from the FSS.

A second example, which brings into play an external system unit, is shown in Fig. 14. Here the order to be processed, derived from operation type 5, is "Write 1 in the BGS at an address whose Y part (vertical coordinate of the 128-by-128 BGS array) is the D code of the order and whose X part (horizontal) is stored in flip-flop group Q". Here the order translator activates four of its outputs, which (a) gate the D code order register flip-flop outputs to the Y half of the bus input, (b) gate the X half flip-flop outputs of FFG Q to the X half of the bus input, (c) ready the bus output to be gated to the BGS address leads and (d) ready
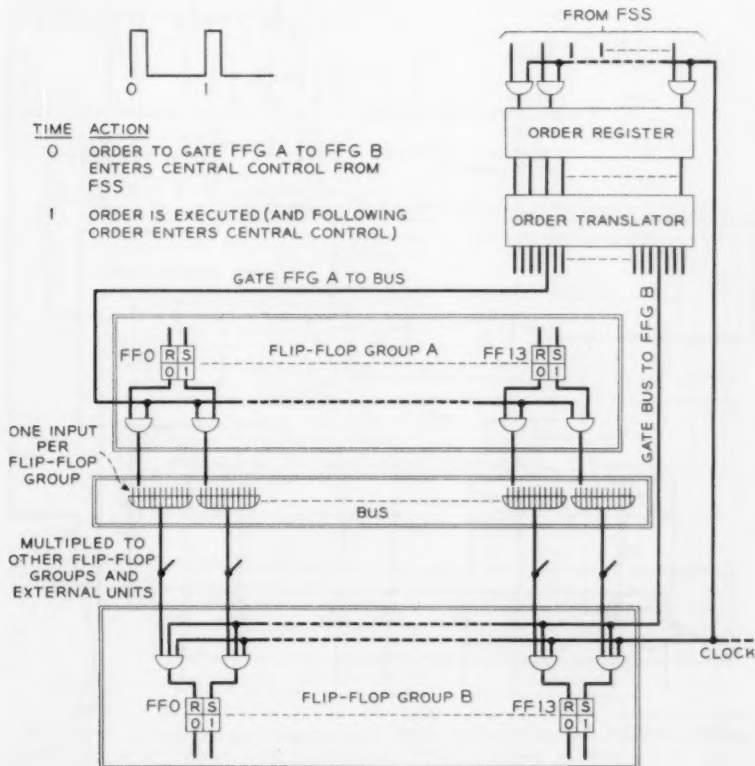
Fig. 13 — Central control circuits for simple nondecision order: "Gate FFG A to FFG B."

a Write 1 signal to be gated to the BGS. When the clock pulse occurs at time 1, this order is executed.

8.2 *Decision Making*

Both of the examples just presented involved nondecision operations; that is, the Gate FFG to FFG and Write 1 in BGS orders can be followed only by the succeeding order in the program, located at the next FSS address. We will now develop the decision-making logic which comes into play on decision (reading or matching) operations when the following
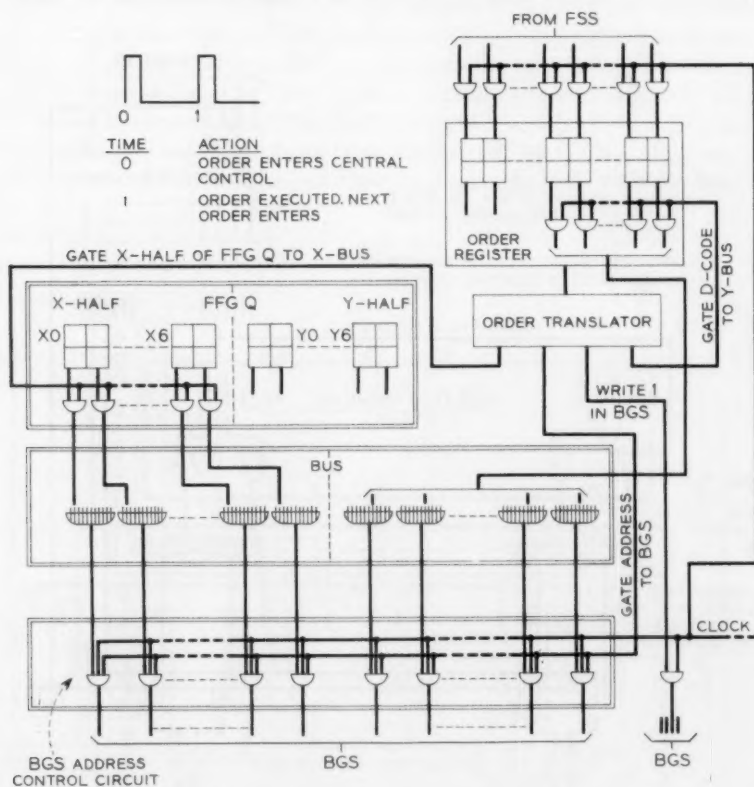

Fig. 14 — Central control circuits for another nondecision order: "Write 1 in BGS at Y address given in D code of order, X address preset in X half of flip-flop group Q."

action is one of two alternatives, and discuss the way the clocking of the system is affected.

A complication arises where decisions are involved because, at the same time an order is executed, it is displaced from the order register by the succeeding order. Since the response to a decision operation occurs *after* the execution of that operation, means must be provided for recovering certain basic information in the order which will determine, when combined with the response, which of two alternative actions is to follow. This point is illustrated in Fig. 15 which shows the timing involved when a Read BGS order (described in detail below the figure) is performed; it should be noted that, if the BGS response is 0, the system continues with
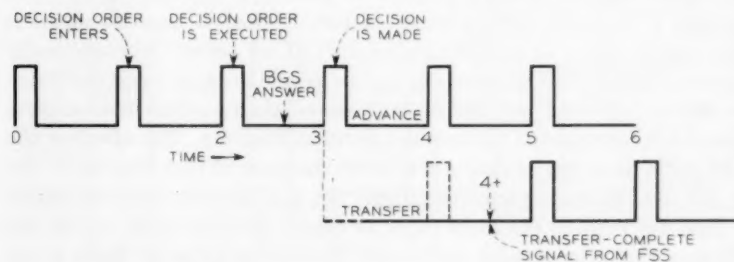
TIME	ACTION
0	NON-DECISION ORDER A ENTERS CENTRAL CONTROL, FSS GOES TO ORDER B
1	ORDER A EXECUTED; (DECISION) ORDER B ENTERS CENTRAL CONTROL; FSS GOES TO ORDER C
2	ORDER B EXECUTED; ORDER C ENTERS CENTRAL CONTROL; FSS GOES TO ORDER D

IF READING IS 0:

TIME	ACTION
3	ORDER C EXECUTED; D ENTERS; FSS TO E
4	ORDER D EXECUTED; E ENTERS; FSS TO F
5	ORDER E EXECUTED; F ENTERS; FSS TO G
ETC	ETC

IF READING IS 1:

TIME	ACTION
3	ORDER C NOT EXECUTED; FSS DIRECTED TO TRANSFER TO ADDRESS STORED IN FLIP-FLOP GROUP T
4+	FSS SIGNALS CENTRAL CONTROL THAT TRANSFER IS COMPLETE
5	ORDER J ENTERS CENTRAL CONTROL; FSS GOES TO ORDER K
6	ORDER J EXECUTED; ORDER K ENTERS CENTRAL CONTROL; FSS GOES TO ORDER L
ETC	ETC

Fig. 15 — Central control timing on decision orders. The decision order, designated order B, is: "Read (and regenerate) the BGS at the address whose Y part is D the code of the order and whose X part is preset in flip-flop group Q. If the reading is 0, continue with the next order (c) in the program; if the reading is 1, transfer to order J, which is another point in the program whose FSS address is (was previously) stored in flip-flop group T."

the next order in the program, whereas if it is 1, a transfer is made. At time 2 the Read BGS order (designated order B) is executed; that is, the Read-Regenerate lead and an address are pulsed to the BGS; also at Time 2, order B is displaced from the order register by order C. The BGS response is returned to central control during the interval between time 2 and time 3. Clearly, if a decision is to be made at time 3 whether or not to transfer, the specification contained in the order as to which sense of BGS response should cause a transfer must be *stored* when the order is displaced at time 2. Then, at time 3, this partial memory of the order may be combined with the BGS response to determine the decision. Here, if the BGS response is 0, order C, which follows order B in the program and which entered the order register at time 2, is executed. If the BGS response is 1, order C is not executed but, instead, a transfer signal is sent together with an address to the FSS. If we assume hypothetically that the FSS requires between one and two cycle times to reach the transfer address and read out the order J stored there, a transfer-complete signal[5] will be returned to central control at time 4+. The effect of the clock pulse occurring at time 4 is inhibited because at that time the transfer has not yet been completed. However, the transfer-complete signal at time 4+ permits the clock pulse at time 5 to bring order J from the FSS into the order register and causes the FSS to go on to order K. At time 6, J is executed, K enters the order register, and the FSS goes on to L, and so forth.

The addition of the order memory and decision logic used in conjunction with decision orders is shown for the previous example in Fig. 16. Here five order translator outputs are activated: one gates the X half of FFG Q to the X bus; a second gates the D code of the order register to the Y bus; a third readies the bus output to be gated to the BGS as an address; a fourth prepares a Read-Regenerate order to be gated to the BGS and a fifth prepares the "Transfer if 1" flip-flop in the order memory to be set. The latter three actions will occur at time 2 (Fig. 15), when the clock pulse passes through the decision logic and exits on the "Execute Present Order" (EPO) lead through gates not shown. Also at time 2, order B will be displaced from the order register by order C, but not before the order memory flip-flop is set to retain the sense of the transfer specified in order B. The BGS response occurs between time 2 and time 3. At time 3, therefore, if the BGS response is 0 an EPO pulse will be generated, causing order C to be executed. (The nature of order C is not specified and the circuits shown should not be assumed to be adequate to process it). If the BGS response is 1, on the other hand, a "Conditional Transfer" (CTR) pulse is generated by the decision logic,
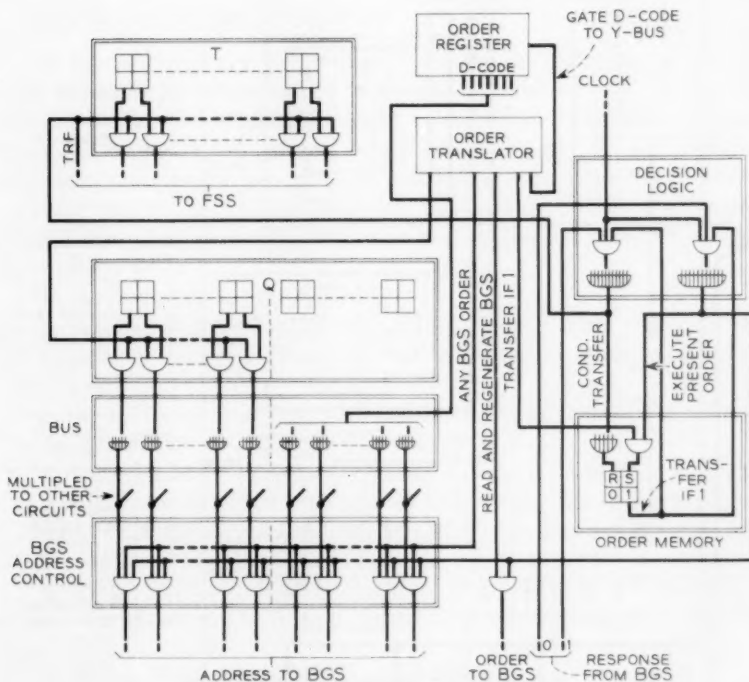
Fig. 16 – Central control circuits for simple decision order: "RY1."

which resets the order memory and causes a transfer signal to be sent to the FSS, together with the address previously stored in FFG T. In general, every clock pulse is converted either to an EPO pulse, which causes whatever order is in the order register to be executed and displaced by the succeeding one, or a CTR pulse, which causes a transfer of the FSS to a new order located at a previously stored address.

8.3 Functional Description of Central Control*

With the basic features of central control design now described, we may consider the more complete functional diagram shown in Fig. 17.

_____

* The material in this section describes in some detail the interrelationships among the several registers and control circuits of the central control. If the reader wishes, he may proceed at this point directly to Section IX (Programming) without loss of continuity.
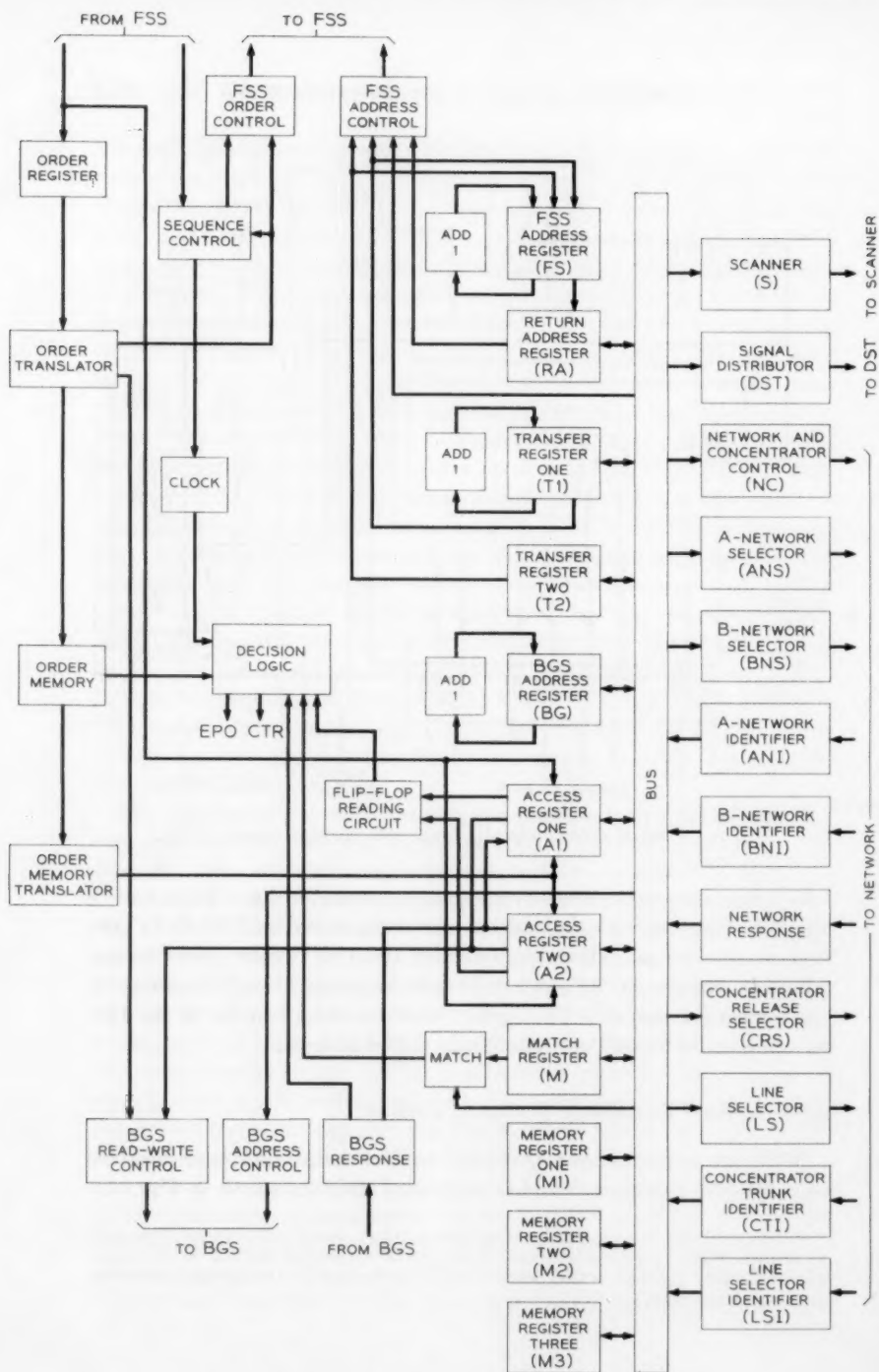
Fig. 17 — Functional block diagram of central control.

8.3.1 *Special-Purpose Registers*

On both sides of the bus in Fig. 17 are shown flip-flop groups which will hereafter be referred to as "registers." On the right of the bus are special-purpose registers which have access either to or from the bus, but not both. Those registers which are used only to communicate control or address information to external system units from central control receive information from the bus output; those used to pass information from external units to central control are connected to the bus input.

In the first class are the following:

1. Scanner Register. Whenever a Read Scanner order is given, this register must already hold the address of the line or trunk whose state is desired.

2. Signal Distributor Register. This holds the address of the relay to be operated or released by the signal distributor.

3. Network and Concentrator Control Register. The code stored in this register instructs the concentrator and distribution network marker circuits what actions to carry out. The addresses involved are stored in other registers.

4. A-Side Network Selector Register. This holds the A-side distribution network address to be used in establishing or releasing a distribution network connection.

5. B-Side Network Selector Register. Same as above for B-side.

6. Concentrator Release Selector Register. This holds the concentrator trunk address when a concentrator connection is released.

7. Line Selector Register. This holds the line address when a line-to-concentrator trunk connection is to be made through the concentrator.

All these registers may be considered functional parts of the external units to which they are connected, since the outputs of their flip-flops directly control the actions in these units. They are made physically a part of central control to avoid the need for transmitting pulses to them from the bus over the distances between central control and external units. With this arrangement, only dc flip-flop outputs must travel over long lead lengths.

The remaining registers shown to the right of the bus in Fig. 17 have access to the bus input, passing information from the network to central control. These are:

8. A-Side Distribution Network Identifier Register. Whenever a connection including an A-side terminal is either established or released this register will store the address of this terminal.

9. B-Side Distribution Network Identifier Register. Same as above for B-side.

10. Concentrator Trunk Identifier Register. This receives a partial address of the concentrator trunk involved in a concentrator connection or release operation.

11. Line Selector Identifier Register. This receives the line address on a concentrator connect operation, or a partial address of a concentrator trunk on a concentrator release operation. The combination of the information in this and the previous register specifies the full address of a concentrator trunk on a release operation, and of a line and concentrator trunk on a connect operation.

12. Network Response Register. The flip-flops in this register receive indications of whether a network action is ended and whether it is successful or unsuccessful.

8.3.2 General-Purpose Registers

To the left of the bus in Fig. 17 are registers which have access both to and from the bus. Some are special-purpose but are available for general-purpose use when their special functions are not required. Others are provided only for general-purpose use. Their functions are described below:

13. Flying Spot Store Address Register. This is included here to facilitate descriptions of other registers, although it has access neither to or from the bus. At all times this register holds the FSS address of the program order awaiting execution in the order register. An Add 1 circuit is associated with the FSS address register so that, as each order is executed and followed by the next order in the program, the address it holds may be incremented. When a transfer occurs, the FSS address to which the transfer is to be made is sent to this register as well as to the FSS.

14. Return Address Register. Frequently a program is used repeatedly and, because of this, is recorded in the FSS as a subroutine. When a program transfers into a subroutine it is often necessary that the program be resumed following the completion of the subroutine, i.e., that the subroutine end with a transfer returning to the program which initiated it. The return address register receives from the flying spot store address register, upon each transfer, the FSS address following the one from which the transfer is made. Thus, a subsequent transfer to the FSS address stored in this register will cause the program from which the original transfer was made to be resumed.

15. Transfer Registers 1 and 2. These hold the FSS address to which a transfer may be made as the result of a decision order. The decision order must specify from which of the transfer registers the address is taken. The reason that two are provided and the function of the Add 1 circuit associated with one of them will be explained in a later section.

16. Barrier Grid Store Address Register. This supplies part or all of a BGS address on all BGS reading and writing orders. As already explained, BGS orders provide at most a seven-bit X or Y address, the remaining seven bits being taken from a flip-flop register. This was the register referred to and which also appeared in the examples of Figs. 14 and 16 as flip-flop group Q. The Add 1 circuit that is associated with it is provided for the Add 1 operation listed in Table V. If a number is to be incremented by 1, it is first placed in the BGS address register and then acted upon by the Add 1 circuit. Thus the register has two special-purpose uses and may be employed for general-purpose use as well.

17. Access Registers 1 and 2. These are the most versatile registers of the central control. They are characterized by the fact that their flip-flops are individually accessible. As indicated by the many paths to and from these registers in Fig. 17, they are used: (a) for communications to and from the bus; (b) when a bit is read from the BGS to be stored in a flip-flop (operation type 7, Table V); (c) when a bit is read from a flip-flop to be stored in the BGS (operation type 8); (d) when a bit is to be read from or written into an individual flip-flop (operation types 2 and 6); (e) to receive translation information from the flying spot store, since individual bits of a translation word must be examined.

18. Matching Register. This is used for the matching operation (type 4). In it is placed a number to be matched with another number stored in any other register with access to the bus. The associated Match circuit compares the two numbers and notifies the decision logic of the result.

19. Memory Registers 1, 2 and 3. These are completely general-purpose registers used most frequently to store program addresses for later use.

The registers having intercommunications by way of the bus may be though of as a rectangular matrix of flip-flops, one dimension being the number of such registers and the other the maximum number of flip-flops per register (14). Not all elements of the matrix are occupied, especially in the case of some special-purpose registers associated with external units where the number of bits required is fewer than 14. Clearly, the flip-flops in such registers must be assigned definite posi-

tions on the bus so that proper communications can be carried on with the remainder of central control.

8.3.3 Control Circuits

On the left side of Fig. 17 are shown the circuit blocks which have control rather than register functions. Among them are the order register, order translator, order memory, decision logic and clock, all of which have already been covered in earlier paragraphs.

Associated with the order memory is an *order memory translator*, whose outputs affect the two access registers. This translator is required for operation type 7, which involves a BGS reading being stored in a flip-flop of one of the access registers. As shown in Table VIb, the five-bit C code identifies this flip-flop. Since the BGS response occurs after the order has been displaced from the order register, the C code must be held in the order memory and translated by the order memory translator to a one-out-of-32 indication to prepare the input logic of the appropriate access register flip-flop for receipt of the BGS response.

The three circuits associated directly with the BGS and shown near the bottom of the figure are the *read-write control, address control* and *response circuit*. The BGS read-write control determines which, if any, of the four BGS orders is to be given (Read and Regenerate, Read and Write 0, Read and Change, Read and Write 1); most often the order translator alone determines this, but, in the case of operation type 8, the state of an access register flip-flop will determine whether a Write 0 or Write 1 signal is given. The BGS address control gates the bus output to the BGS address leads as shown on Figs. 14 and 16. The BGS response circuit output is connected to the decision logic for determining whether or not to transfer on BGS decision operations, and to the access registers for storing a bit read from the BGS into a flip-flop for later use.

The *flip-flop reading circuit* shown in Fig. 17 is essentially a large OR gate which combines the outputs of all the flip-flops of central control which can be individually read (using operation type 2). During the processing of a Read Flip-Flop order, the seven-bit flip-flop address is converted by the order translator to a signal which permits only the output of the corresponding flip-flop to be transmitted into the flip-flop reading OR gate. In this way the state of this flip-flop appears at the output of the OR gate.

The *flying spot store address control* at the top of Fig. 17 receives a FSS address from one of a number of possible sources and transmits it to the FSS when a transfer is made. Included among these sources are

the two transfer registers, the return address register and the bus. The FSS address of a transfer also is stored in the FSS address register of central control, as already explained.

The *flying spot store order control* generates the orders to the FSS. Except for a few special situations, whenever the decision logic converts a clock pulse to an Execute Present Order pulse, an Advance pulse is sent to the FSS, which causes the succeeding order to be gated from an output register in the FSS to the order register in central control, and then causes this same output register to be filled with the next order. When the decision logic converts a clock pulse to a Conditional Transfer pulse, this circuit sends a Transfer signal to the FSS and the FSS accepts a new address from central control rather than stepping along, as it does in the case of an Advance.

The last block yet to be described in Fig. 17 is the *sequence control*. This is the only basically sequential circuit in the central control. It functions whenever control of the system is temporarily taken away from the program or whenever completely synchronous clock control gives way to asynchronous control actions. The two major uses of this circuit are in the cases of transfers and FSS translations.

The time required to complete a transfer is somewhat variable over a small range because of the characteristics of the FSS.[5] The sequence control therefore acts to inhibit clock pulses following the one on which a transfer occurs from affecting the central control until it receives a Transfer-Complete signal from the FSS. It causes the clock pulse following this signal to gate the order at the transfer location from the FSS to the order register and then returns normal synchronous control to the clock.

In the case of a FSS translation, a transfer is made to an address where a translation word, not an order, is to be read out and stored directly in the access registers. Clearly, a separate control circuit is necessary to accomplish this, since the FSS cannot be used simultaneously to control the system with an order and to act as a source of translation information. Here, the sequence control steers the FSS output into the access registers rather than the order register. It then causes a second transfer to the address held by the return address register, causing control to be returned to the program order following the one which initiated the FSS translation.

This completes a functional description of central control. All functions are realized through the use of diode AND and OR gates, flip-flops and inverters. Added to this basic logic are amplifiers, emitter-followers and other supporting circuits. A central control like the one

described would contain on the order of 2000 transistors and 15,000 diodes.

IX. PROGRAMMING

This final section deals with the writing of the programs which will be recorded on the photograph plates of the flying spot store. A program is a group of orders, each encoded into a 17-bit word, which specifies in complete detail what the system is to do. From the programmer's viewpoint, the nature of the central control and other circuit action occurring in the execution of the order is of little concern. It is essential that the programmer know only the action taken as a result of each order, and, in all cases, this is specified by its definition.

This section of the article will, therefore, have little direct relation to the previous one. Its purpose is to explain how electronic switching system programs are composed and, in the process, to illustrate a few of the problems confronting the programmer.

9.1 *A Detailed Order Structure*

Before programs can be written there must exist a complete list of available program orders, each precisely defined. The basis for such a list was established earlier and is shown in Table VIb. It is now appropriate to derive from this list of operation types a set of program orders. Considerations related to those discussed earlier are used in arriving at variations of the operation types which will prove to be useful orders. The patterns of system actions depicted on the sequence charts are especially helpful in forming judgments as to the potential utility of a proposed order. Ease of circuit implementation is also a contributing factor.

After a tentative list of orders is composed and some programs are written, orders which are found to have infrequent application and can be replaced by combinations of other orders are deleted; others may be added as a result of discoveries of useful applications by programmers. A stable order structure eventually emerges.

The order structure to be used in this article is shown in Table VII. The listing includes 38 orders, 16 decision and 22 nondecision. For each order is shown (a) a symbolic designation having mnemonic significance, (b) the operation type in Table VIb from which it is derived, (c) its numerical (ABCD) coding in decimal form and (d) an abbreviated but precise definition.

TABLE VII

No.	Symbol	Operation Type (from Table VIb)	A, B	C	D	Description
						Decision Orders
1	RY-	1	*	0	—	Read and regenerate BGS at Y specified by D (X given by BGX).
2	RX-	1	*	1	—	Read and regenerate BGS at X specified by D (Y given by BGY).
3	EY-	1	*	2	—	Read and erase BGS at Y specified by D (X given by BGX).
4	EX-	1	*	3	—	Read and erase BGS at X specified by D (Y given by BGY).
5	CY-	1	*	4	—	Read and change BGS at Y specified by D (X given by BGX). Suffix refers to reading before change.
6	CX-	1	*	5	—	Read and change BGS at X specified by D (Y given by BGY). Suffix refers to reading before change.
7	RP-	1†	*	6	—	Read and regenerate the BGS at address stored in BG.
8	RYB-	1	*	7	—	Read and regenerate at Y specified by D (X given by BGX). Store Y in BGY.
9	RXB-	1	*	8	—	Read and regenerate at X specified by D (Y given by BGY). Store X in BGX.
10	EYB-	1	*	9	—	Read and erase at Y specified by D (X given by BGX). Store Y in BGY.
11	EXB-	1	*	10	—	Read and erase at X specified by D (Y given by BGY). Store X in BGX.
12	RFF-	2	*	11	—	Read miscellaneous flip-flop specified by D.
13	RS-	3	*	12	—	Read the scanner at address preset in S.
14	MY-	4†	*	13	—	Match the Y bits of M with the Y bits of the FFG specified by D. Match = 1; mismatch = 0.
15	MX-	4†	*	14	—	Match the X bits of M with the X bits of the FFG specified by D. Match = 1; mismatch = 0.
16	MB-	4	*	15	—	Match the contents of M with the contents of the FFG specified by D. Match = 1; mismatch = 0.
						Decision Order Modifier Suffixes
—	-0	—	0, 0	—	—	If 0 is read, take next order from address given by T1.
—	-1	—	1, 0	—	—	If 1 is read, take next order from address given by T1.
—	-02	—	0, 2	—	—	If 0 is read, take next order from address given by T2.
—	-12	—	1, 2	—	—	If 1 is read, take next order from address given by T2.
—	-0A	—	0, 1	—	—	Add 1 to Y part of contents of T1. Then, if 0 is read, take next order from address given by T1.
—	-1A	—	1, 1	—	—	Add 1 to Y part of contents of T1. Then, if 1 is read, take next order from address given by T1.

TABLE VII — *Continued*

No.	Symbol	Operation Type (from Table VIb)	A, B	C	D	Description
						Nondecision Orders
17	W0Y	5	7, 3	0	—	Write a 0 on the BGS at the Y address specified by D (X given by BGX).
18	W1Y	5	7, 3	1	—	Write a 1 on the BGS at the Y address specified by D (X given by BGX).
19	W0X	5	7, 3	2	—	Write a 0 on the BGS at the X address specified by D (Y given by BGY).
20	W1X	5	7, 3	3	—	Write a 1 on the BGS at the X address specified by D (Y given by BGY).
21	W0P	5†	7, 3	4	—	Write a 0 on the BGS at the address contained in BG.
22	W1P	5†	7, 3	5	—	Write a 1 on the BGS at the address contained in BG.
23	W0FF	6	7, 3	6	—	Write a 0 in the miscellaneous flip-flop specified by D.
24	W1FF	6	7, 3	7	—	Write a 1 in the miscellaneous flip-flop specified by D.
25	RYFA	7	3, 0	—	—	Read and regenerate the BGS at the Y address specified by D, and transport the bit to the access flip-flop specified by C (X given by BGX).
26	WFAY	8	3, 1	—	—	Transport the contents of the access flip-flop specified by C to the BGS at the Y address specified by D (X given by BGX).
27	ST1	9	4, -	—	—	Set up transfer register 1 to the number specified by B, C, D.
28	SA1	9	5, -	—	—	Set up access register 1 to the number specified by B, C, D.
29	SA2	9	6, -	—	—	Set up access register 2 to the number specified by B, C, D.
30	SY	9†	7, 3	10	—	Set up BGY to the number specified by D.
31	SX	9†	7, 3	11	—	Set up BGX to the number specified by D.
32	G	10	3, 3	—	—	Gate the contents of the FFG specified by C to the FFG specified by D.
33	T	11a	2, -	—	—	Transfer to the address specified by B, C, D.
34	TFG	11b	7, 3	12	—	Transfer to the address contained in the FFG specified by D.
35	AY	12	7, 3	13	—	Add 1 to BGY.
36	AX	12	7, 3	14	—	Add 1 to BGX.
37	RGY	13	7, 3	15	—	Regenerate the BGS at the Y address specified by D (X given by BGX).
38	RGX	13	7, 3	16	—	Regenerate the BGS at the X address specified by D (Y given by BGY).

* Modifier code.
† Size of address not same as specified in Table VI.

A decision order must be suffixed with a modifier indicating which sense of response should cause a transfer. There are six modifiers: (a) "0" suffixed to a decision order (e.g., RY0) means transfer on a 0 response to the FSS address stored in transfer register 1; (b) the "1" suffix causes the transfer on a 1 response; (c) and (d) the 02 and 12 suffixes are the same as those above, respectively, except that the transfer address is taken from transfer register 2; (e) and (f) the 0A and 1A modifiers are similar to 0 and 1 except that the FSS Y address held by transfer register 1 is incremented by 1, the purpose of this to be illustrated by a later example. Thus, by adding one of the six modifiers to a decision order it is possible to cause a transfer on either a 0 or 1 response to the FSS address stored in either transfer register 1 or 2, and, if transfer register 1, either changing or not changing the Y part of the address stored there.

The 16 decision orders are derived from the first four operation types. The RY-* order causes the BGS to be read and regenerated at the Y address specified by the D code and the X address preset in the X half of the barrier grid register of central control. The RX- does the same, except that the D code is the X address and Y is preset. The EY- and EX- orders are similar, but the BGS is read and erased instead of regenerated. The CY- and CX- cause the BGS to be read and changed; here, for decision purposes, the response is assumed to be the reading before the change occurs. The RYB-, RXB-, EYB- and EXB- are the same as RY-, RX-, EY- and EX- except that the D code is gated to the Y or X half of the barrier grid register in central control as well as to the BGS as a Y or X address. This is provided for those cases where it is useful to keep track in the barrier grid register (BG) of the BGS address last visited. The RP- order causes the BGS to be read and regenerated with the entire address preset in BG, the order itself containing no address.

The MY- order causes the contents of the Y half of the register specified by the D code to be matched against the contents of the Y half of the match register. The 1 suffix is used for the "match" response and the 0 suffix for "mismatch". MX- accomplishes the same for the X halves and MB- for both halves together, or the entire contents of both registers.

The RFF- and RS- orders are the only ones derived from operation types 2 and 3 and their definitions are self-explanatory.

The 22 nondecision orders are derived from operation types 5 through

---

* The hyphen stands for a modifier, which must be attached.

13 of Table VI, and their descriptions in Table VII require no amplification.

9.2 *The Register Matrix*

Combinations of the program orders described above are used to effect all system actions, many of which involve the central control together with some other system unit. Because the major system components (Fig. 1) differ in function and size, a "numbering plan" is formed which ties them together in proper association. In central control, the numbering plan is reflected to some extent in the relationship of each flip-flop register to the bus. Table VIII shows the bus having 14 positions divided into X and Y coordinates. The general-purpose registers occupy the full 14 positions, whereas the registers associated with other system units are tailored, both with respect to size and placement on the bus, to the numbering plan.

If a 12-bit scanner address were to be taken from the BGS and placed in the scanner register via access register 1, for example, the 12 bits

TABLE VIII

Register (Abbreviations shown in Fig. 17)	X Position							Y Position						
	X_6	X_5	X_4	X_3	X_2	X_1	X_0	Y_6	Y_5	Y_4	Y_3	Y_2	Y_1	Y_0
A1	13	12	11	10	9	8	7	6	5	4	3	2	1	0
A2	27	26	25	24	23	22	21	20	19	18	17	16	15	14
BG	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
BGX*	≡	≡	≡	≡	≡	≡	≡							
BGY*								≡	≡	≡	≡	≡	≡	≡
T1	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
T2	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
RA	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
M	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
M1	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
M2	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
M3	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
S			≡	≡	≡	≡	≡							
DST			≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
ANS							≡	≡	≡	≡	≡	≡	≡	≡
BNS							≡	≡	≡	≡	≡	≡	≡	≡
CRS							≡	≡	≡	≡	≡	≡	≡	≡
LS			≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
ANI								≡	≡	≡	≡	≡	≡	≡
BNI								≡	≡	≡	≡	≡	≡	≡
CTI														≡
LSI			≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
NC			≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡	≡
NR								≡	≡	≡	≡	≡	≡	≡

* BGX and BGY together comprise BG but are separately addressable.

would have to be placed from the BGS into access flip-flops 0, 1, 2, 3, 4, 5, 7, 8, 9, 10, 11 and 12. Access register 1 could then be gated to the scanner register, since these flip-flops occupy the same bus positions as the 12 flip-flops of the scanner register.

9.3 *Examples of the Uses of Program Orders*

In this section several examples of the uses of program orders are given. Each example contains a sequence to be programmed, any BGS layout involved in the sequence and one or more programs fulfilling the requirements of the sequence. In the examples the symbolic languages for sequences and programs already presented will be employed. Frequent reference to Table VII will be helpful.

Example 1 — Fig. 18

The sequence in this example requires the writing in two BGS locations having different Y addresses (A, T) and a common X address (R1). The corresponding program contains three orders. The order SX = R1 is used to place BGS X-address R1 in the BG register of central control (see Table VIII); this is necessary since orders to read or write in the BGS contain at most an X *or* Y address, the other half being preset in BGY or BGX, respectively. The orders W1Y = A and W0Y = T are used to write 1 in R1-A and 0 in R1-T, A and T being the Y addresses specified in these two orders with X preset as a result of the previous order. The = sign in a program order separates the operation part of the order from the address(es) at which the operation is carried out.
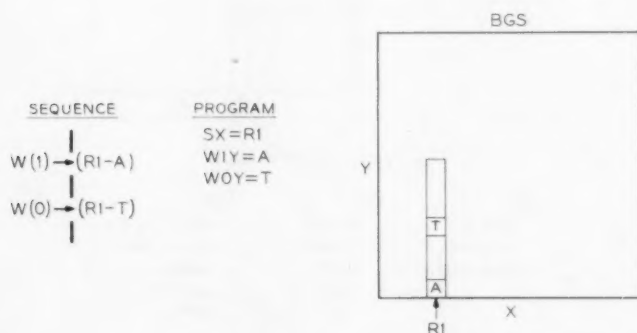
Fig. 18 — Programming example 1.

Example 2 — Fig. 19

The sequence here requires writing 1 at two BGS locations having a common Y address (T) and different X addresses (V1, V2). Three alternative programs are shown.

Program (a) is analogous to the previous example, except that X and Y are interchanged and 1 is written at both BGS locations.

Program (b) begins with the orders SY = T and SX = V1, placing the entire BGS address of V1-T in the BG register. The order W1P causes a 1 to be written at this address. Then SX = V2 and W1P cause a 1 to be written at V2-T.

Program (c) is similar to (b) except that the BGS address V1-T is placed in the BG register by first placing it in access register 1 (A1) and then gating the contents of A1 to BG. Thus the orders SA1 = V1, T (which places V1 in the X half and T in the Y half of A1) and G = A1, BG accomplish this. The remaining three orders are identical to those of (b).

Program (a) requires the fewest orders (three) and would therefore be used. Programs (b) and (c) are shown to introduce the use of the W1P, SA1 and G orders.

BGS

SEQUENCE

W(1) → (V1-T)

W(1) → (V2-T)

Y

T T

V1 V2   X

ALTERNATIVE PROGRAMS

(a)	(b)	(c)
SY=T	SY=T	SA1=V1,T
W1X=V1	SX=V1	G=A1,BG
W1X=V2	W1P	W1P
	SX=V2	SX=V2
	W1P	W1P
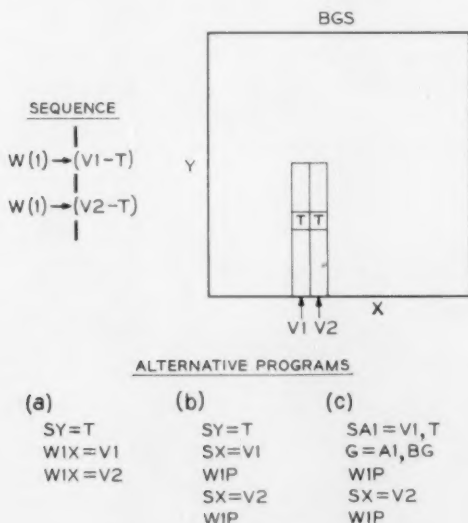
Fig. 19 — Programming example 2.

Example 3 — Fig. 20

This sequence involves a decision based on the reading of BGS location NR-A. If the reading of NR-A is 1, a 1 must be written in NR-RW; if it is 0, NR-A must be changed to 1.

In program (a) the decision order RY1 = A is preceded by ST1 = r, which places in transfer register 1 the address r of the program to be transferred to if the reading of NR-A is 1. In program (b) the transfer is made instead when the reading of NR-A is 0 and the two writing orders following it are interchanged with respect to program (a). The choice between these two programs in this case would be to make the transfer the lower-probability event, since it consumes more time than advancing to the next program order.

Program (c) is similar to (a) with X and Y interchanged. In this case, the group of orders SY = A, ST1 = r, RX1 = NR have exactly the same over-all effect as the corresponding orders of (a). In (c), however,
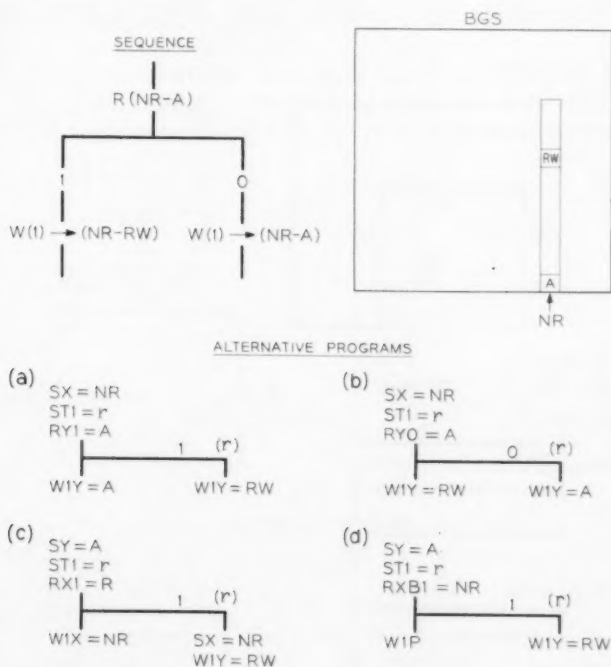
Fig. 20 — Programming example 3.

an additional order (SX = NR) is necessary before a 1 can be written in RW, since the X address NR must be placed in the X half of the BG register before W1Y = RW will have the desired effect.

The additional order is avoided in (d), where the RXB1 order is used in place of RX1. The order RXB1 performs the reading function of RX1, but also causes the address part (NR) of the order to be gated to the X half of the BG register, making unnecessary the order SX = NR used in (c). Also, W1X = NR in (c) is replaced by W1P in (d), since the combination of SY = A and RXB1 = NR result in the full address of NR-A being in the BG register.

Example 4 — Fig. 21

The sequence in this example involves a hunt among a group of BGS registers R1, R2, $\cdots$ RN for one in which the A location reads 0. When such is found a 1 is written in the A location and the program reaches
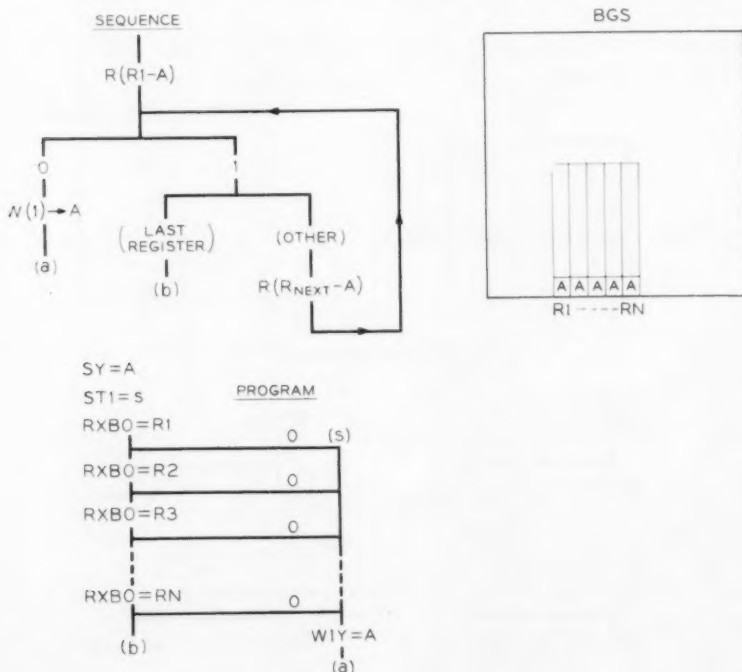
Fig. 21 — Programming example 4.

point a. If none is found the program reaches point b. Frequent use of the RXB order occurs in this type of program.

Example 5 — Fig. 22

This sequence illustrates the use of the Read-and-Erase (EY and EXB) orders. The sequence requires that a reading be made of NR-A; if the reading is 1, NR-A should be written to 0. Programs (a) and (b) are otherwise similar to programs already described.

Example 6 — Fig. 23

This example illustrates the use of the Read-and-Change (CY) order. The sequence requires that a counter (PC) containing two bits (PCA, PCB) be incremented by 1 followed by a 1 being written into OR1-TR. The incrementing is done by first changing the value of the least significant bit (PCA). If the value of PCA is changed from 0 to 1, the process is finished; if PCA is changed from 1 to 0, the value of PCB is changed and the process is finished.

Program (a) accomplishes this using types of orders already described in previous examples. After setting the X half of the BG register to OR1 and transfer register 1 to v, the PCA bit is read (its Y address
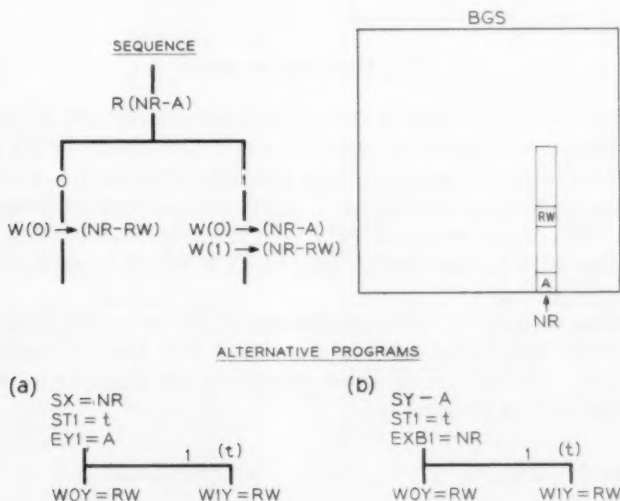
Fig. 22 — Programming example 5.
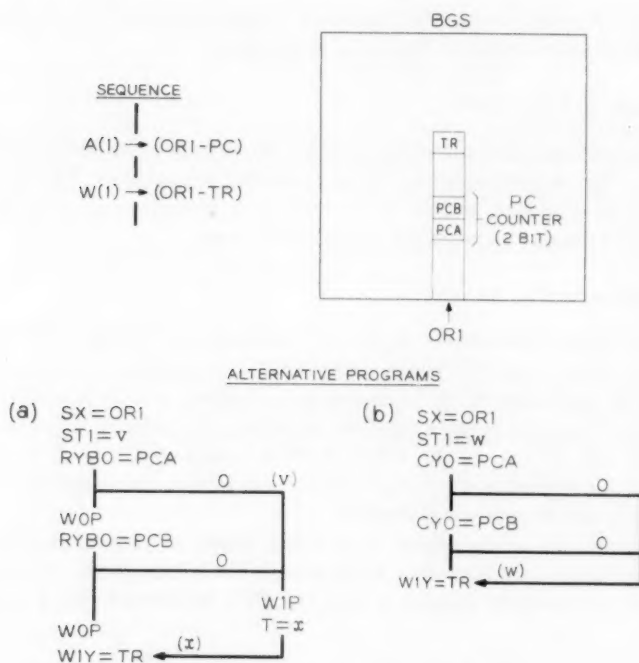
Fig. 23 — Programming example 6.

being gated to the Y half of the BG register since RYB0 is used). If the reading is 0, a transfer is made to v and a 1 is written in PCA by the W1P order; since the process is then finished, a transfer is made by the $T = x$ order to program location x, where the sequence continues with W1Y = TR. If the reading of PCA were 1, on the other hand, it would be written to 0 by the W0P order and PCB would be read with the pattern repeating.

Whereas program (a) contains nine orders, the use of the CY order in program (b) reduces the number of orders to five. The CY0 order both reads a bit and changes its value, combining the functions of RYB0, W0P and W1P in program (a).

Example 7 — Fig. 24

The sequence of this example calls for three successive decisions based on the reading of three BGS locations (NR-P, Q and R). Since

presumably different actions take place on each possible leg of the sequence (a, b, c, d), a different program location must be provided for each.

Program (a) contains seven orders and fulfills the requirements of the sequence in a straightforward manner. Note that a separate ST1 order must be used preceding each RY0 order to provide the different program locations a, b and c.

In program (b) the repeated use of ST1 is eliminated by using the RY0A order (refer to Table VII–decision order modifiers). This order performs the function of RY0 but in addition increments the address in transfer register 1. Thus, program location b becomes related to
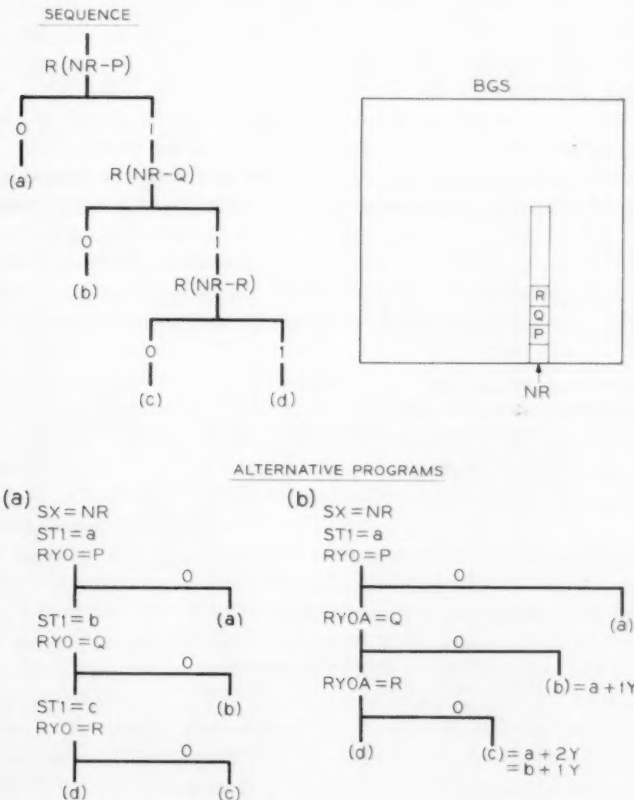
Fig. 24 — Programming example 7.

program location a by a difference of 1 in the Y coordinate, and the two may be used as starting points for completely different and unrelated programs. The same applies to program location c.

Example 8 — Fig. 25

This example illustrates the programming of a simple subroutine. A subroutine may be thought of as a program required under many and varied circumstances but recorded only once to conserve program space. Thus a subroutine may be transferred into from a number of sources and must, upon completion, cause a transfer back to a place associated with the source involved.

The sequence shown should be considered extracted from a larger sequence and requires that a 1 be written in OR–A1 of the BGS. But there are several OR's, and the X address of the one desired is recorded elsewhere in the BGS, in NR-ORX1 to NR-ORX7. These seven bits must be extracted and transported to the X half of the BG register to steer action to the appropriate one of the several OR's. Since this action occurs repeatedly in different parts of the program, the process of extracting the seven bits from NR is made a subroutine.
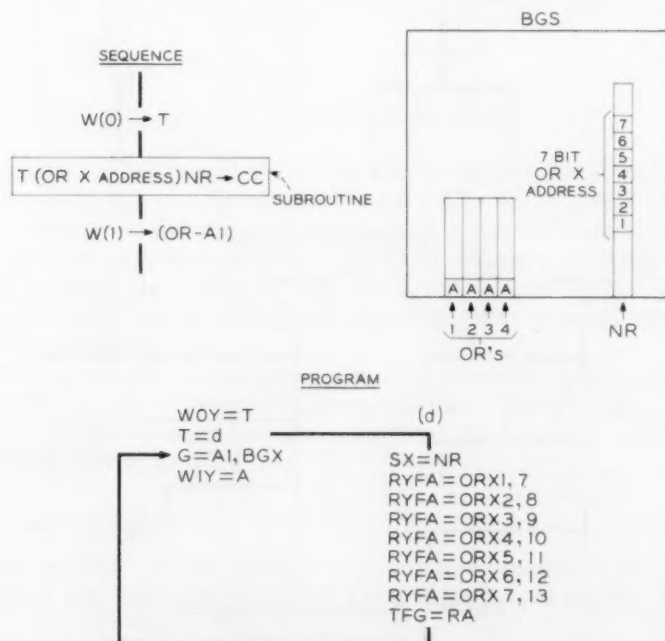
Fig. 25 — Programming example 8.

The program first transfers to the subroutine located at d ($T = d$). The subroutine begins with $SX = NR$ to confine succeeding BGS operations to the proper X address. This is followed by a series of seven RYFA orders, each reading one of the seven bits from the BGS to a flip-flop in access register 1; the address part of these orders contains both the BGS Y address of the bit being read and the identity of the access register flip-flop into which the bit is placed. It should be noted from Table VIII that the bus positions of the seven flip-flops chosen coincide with those of the X half of the BG register. The subroutine ends with $TFG = RA$, which causes a transfer back to the return address, one address beyond that from which the transfer to the subroutine was made, as stored in the return address register. The order $G = A1, BGX$ then gates the seven bits from flip-flops of access register 1 to corresponding flip-flops of the BG register, and $W1Y = OR\text{-}A1$ causes a 1 to be written in the A1 location of the proper one of the OR's.

9.4 *Programming a Larger Sequence*

In this section some of the techniques introduced in the foregoing examples are brought together in the programming of the larger sequence shown in Fig. 26*; the associated BGS layout is given in Fig. 27. A program satisfying the requirements of the sequence is given in Fig. 28. Program locations (flying spot store addresses) a, b, c, d, e, f are shown both in the program and on the sequence chart to relate corresponding portions of the two.

Beginning at point a in the sequence, a reading must be made of OR1-A1, a location in the BGS. If it is 0, OR2-A1 is read, and so forth through all the OR's in the BGS. If a 1 is read, on the other hand, the sequence requires 1 to be added to AIT and then a reading to be made of AIT. If it is 1, OR(NEXT)-A1 is read as before; if it is 0 (point b) a reading is made of LL.

The program for this portion of the sequence is shown in Fig. 28 at location a. A pattern of alternation of $RXB0 = OR\text{-}$ and $CY12 = AIT$ orders is evident. The RXB0 orders perform the function of reading A1 for each OR; each RXB0 is preceded by ST1, so that if the reading is 0 a transfer is made to repeat the process for the next OR. If the reading on RXB0 is 1, $CY12 = AIT$ is performed, adding 1 to the AIT counter. A transfer to the address in transfer register 2 (point b) will then take place if AIT was 1 and was changed to 0; otherwise, the next OR will be treated. The result of the process to this point is that a transfer to b will occur for each OR satisfying the conditions $A1 = 1$, $AIT = 1$ (before the addition of 1, making $AIT = 0$) as required by the sequence.

---

* The particular system function performed by the sequence is of no concern in this section and will not be described.

The correspondence between the remainder of the program and the requirements of the sequence is left to the reader.

9.5 *A Typical Programming Problem*

Previous examples have clearly shown that a number of different programs can be written which satisfy the requirements of a given sequence. The programming problem is therefore not primarily one of arriving at a correct program as much as arriving at one which is both correct and optimal. In programming a real-time system like the one described in this paper, two minimization criteria apply: program space (number of orders) and program time (time consumed per unit of real
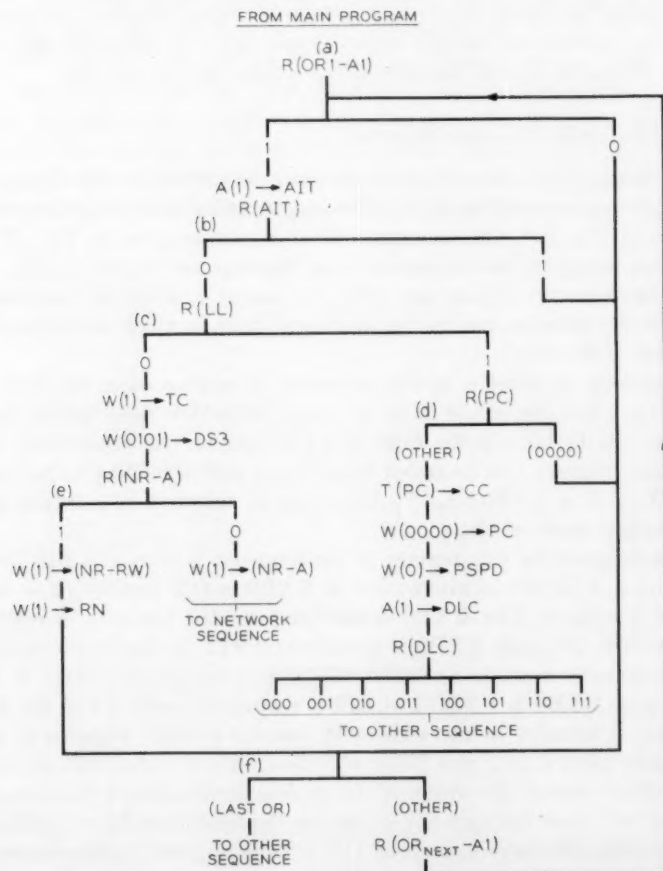
Fig. 26 — A sequence chart to be programmed: Abandoned and interdigital timing (AIT).

time). The optimal program is the one which minimizes some elusive function of these two variables.

In most programming problems the two effects are opposite; i.e., program space reduction is usually accompanied by an increase of program time, and vice versa. A simple example of this is illustrated in Fig. 29. Program (a) contains 65 orders and consumes 65 cycle times; program (b) contains only 11 orders but consumes 259 cycle times.

In general, the programmer combines qualitative and quantitative judgments in deciding the relative merits of alternative courses of action in this and similar situations.

9.6 *The Assembly of Programs*

The writing of programs in symbolic mnemonic form has the valuable advantage of detaching the programmer from unwieldy numerical coding details which would tend to increase the number of errors made. On the other hand, symbolic programs necessitate a process of conversion to the binary form in which programs are actually recorded in the flying spot store.

The latter process, referred to as assembly, is mechanized by the use of a digital computer. Into the computer are placed a symbolic program,
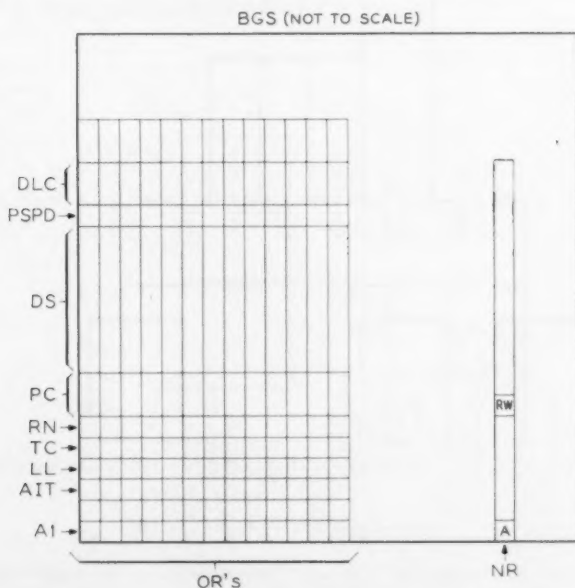
Fig. 27 — Associated BGS layout for AIT scan program.

Fig. 28 — Scan program for AIT.
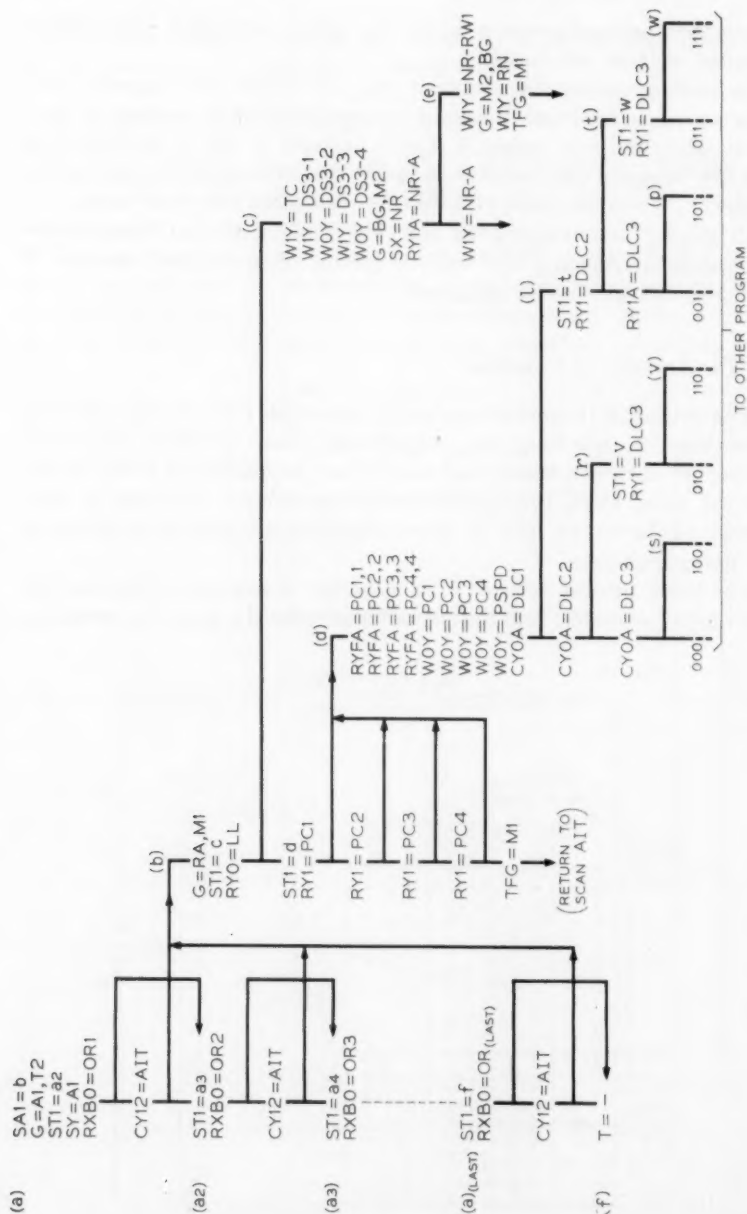
a series of tables specifying the conversions from symbols to binary numbers of the operation-parts and address-parts of program orders, and a set of rules in the form of a computer program for applying the conversions to the symbolic program. Out of the computer comes the binary encoded result ready to control the exposure of the photographic plates of the flying spot store.

9.7 *Simulation*

In the testing of a large stored program system, there is the serious problem of initially distinguishing program errors from equipment troubles. In the case of the experimental electronic switching system an attempt was made to attain an error-free program, so that any difficulties encountered later in laboratory tests of the system could be definitely attributed to equipment.

To accomplish this a large general-purpose digital computer was used to simulate the physical system while the latter was still in its final stage of design. The computer was programmed to image all the memory of the system, accept program orders written in electronic switching system language and interpret and carry out these orders on the system image. In this way, "telephone calls" were processed in the computer and information was printed to indicate the success or failure
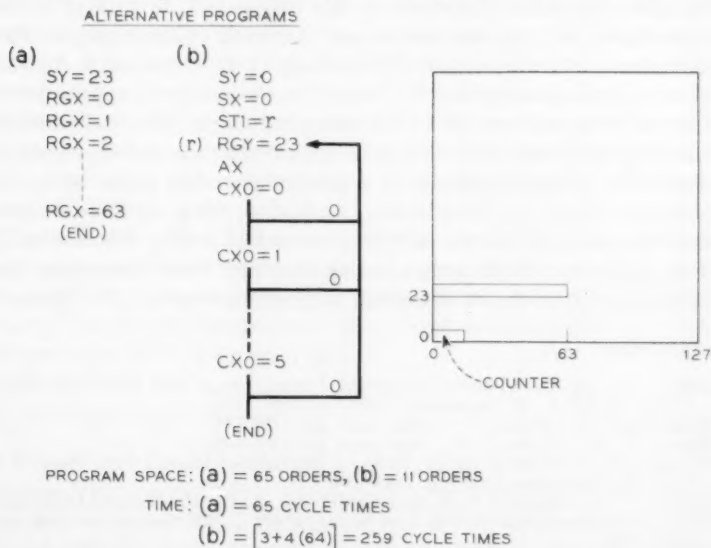
Fig. 29 — Time vs. program space. Sequence: "Regenerate the 64 BGS locations in the left half of row 23. Space for a six-bit counter is available if desired."

of the calls; if the latter, additional information was made available aiding the location of the part(s) of the electronic switching system program causing the failure. After corrections were made based on this information, the process was repeated until the computer reported perfect operation of the program.

X. SUMMARY

This article describes the design of a stored-program electronic switching system from a functional viewpoint. Included are: (1) a description of the major functional units of the system, (2) the introduction of a language and form to describe the sequence of system actions, (3) some examples of how these sequences are created, (4) a presentation of the inter-relationships among sequences in the form of a general pattern of system operation, (5) the development and encoding of an order structure for a stored program, (6) some basic considerations in the design of circuit logic to implement this order structure, (7) a functional description of the organization of this circuit logic, (8) several examples of how programs are written and (9) brief discussions of the processes of program assembly and system simulation.

XI. ACKNOWLEDGMENT

A large number of people share responsibility for the philosophies, techniques and results described in this article and, in writing it, the authors have tied together and related the work of these people. Particular attention is drawn to W. A. Budlong, G. G. Drew and J. A. Harr because of their substantial contributions to the successful application of the stored program concept to telephone switching. The basic method of handling telephone call functions as described by the system sequences in Part One is an outgrowth of engineering studies made by C. E. Brooks and Miss G. E. Markthaler. In addition, the generalized description of the mode of system operation embodied in Fig. 8 is due to H. Ghiron, and some of the programming examples were taken from material prepared for a course of training in programming by J. W. Howard.

REFERENCES

1. Joel, A. E., An Experimental Switching System Using New Electronic Techniques, B.S.T.J., **37**, September 1958, pp. 1091–1124.
2. Feldman, T. and Rieke, J. W., this issue, pp. 1421–1453.
3. Feiner, A. and Goeller, L. F., this issue, pp. 1383–1403.
4. Greenwood, T. S. and Staehler, R. E., A High-Speed Barrier Grid Store, B.S.T.J., **37**, September 1958, pp. 1195–1219.
5. Hoover, C. W., Staehler, R. E. and Ketchledge, R. W., Fundamental Concepts in the Design of the Flying Spot Store, B.S.T.J., **37**, September 1958, pp. 1161–1194.
6. Freimanis, L., this issue, pp. 1405–1419.
7. Yokelson, B. J., Cagle, W. B. and Underwood, M. D., Semiconductor Design Philosophy for the Central Control of an Electronic Switching System, B.S.T.J., **37**, September 1958, pp. 1125–1160.

A High-Speed Line Scanner for Use in an Electronic Switching System

By A. FEINER and L. F. GOELLER, JR.

A high-speed multiposition electronic switch has been developed to act as a time-division information-gathering device for the experimental electronic switching system. This switch, called the scanner, provides central control with random access to all lines and trunks terminating in the central office and permits sampling their states for the reception of supervisory and dial pulse signals.

I. INTRODUCTION

An earlier article[1] on the experimental electronic switching system described the role of a scanner in an electronic telephone office. The scanner, which is the real-time data input circuit for the system, can be thought of as a multiposition electronic switch. Directed by a binary-coded address from the central control, the scanner samples the voltage present at any one of its input terminals and transmits a quantized one or zero signal to central control. All lines and trunks are examined periodically at a rate fast enough to detect dial pulses as well as super-visory changes in state.

In addition to lines and trunks, the scanner connects to a number of test points in other functional units of the electronic switching system to simplify automatic testing for preventative or corrective mainte-nance. The place of the scanner within the electronic switching system is shown in the block diagram, Fig. 1.

1.1 *Line and Trunk Circuits*

The scanner is a voltage-operated device and as such draws very little current from the circuits it samples. Scanner voltages are developed in line and trunk circuits as shown in Fig. 2. Line circuits, which are asso-ciated with the great majority of scanner inputs, operate in a straight-forward manner. Very little current flows from the −50-volt central
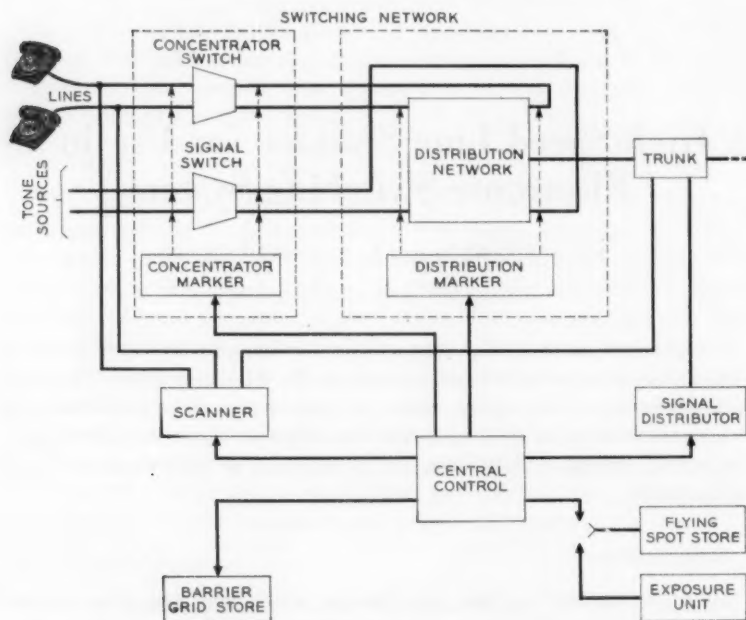
Fig. 1 — Block diagram of the experimental electronic switching system.

office battery when the customer's phone is on hook. Thus, point A is at ground and point B is at −50 volts. When the customer picks up his phone, the loop is closed through the subset and current flows through R_A and R_B, driving A negative with respect to ground and B positive with respect to −50 volts. The low-current subset designed for use in the experimental electronic switching system requires the central office to provide a dc termination of 2300 ohms. Thus, the terminating resistors offer a convenient source of signal voltage to operate the scanner input circuits.

Trunk circuits differ considerably from line circuits and require modified scanner inputs. The trunk circuit shown in part in Fig. 2(b) (as well as the various test points throughout the system which must also be scanned) transmits a signal of +15 volts or ground to the scanner when transistor Q_1 is cut off or saturated.

As long as negligible current flows through the ring side of the trunk, the voltage across R_2 is less than the forward drop across D_2. Thus, very little current flows through the base-emitter junction of Q_1, D_1 and R_2;

Q_1 is cut off and the scanner sees $+15$ volts through R_3. When the opera-
tor at the 3 CL switchboard seizes the trunk, however, she connects the
ring conductor to -24 volts. Current flows from ground through R_2,
the voltage across R_2 exceeds the forward drop across D_2 and the emitter
of Q_1 is driven negative with respect to its base. The transistor is sat-
urated and a voltage very close to ground is transmitted to the scanner.
When the operator at the switchboard dials, the connection between the
ring conductor and the -24 volt battery is interrupted at the dial pulse
rate; the collector voltage at Q_1 follows the dial pulses; L_1, C_1 and D_1
protect the circuit from noise and foreign potentials.

In the scanner, voltages from a line or trunk are used to bias a diode.
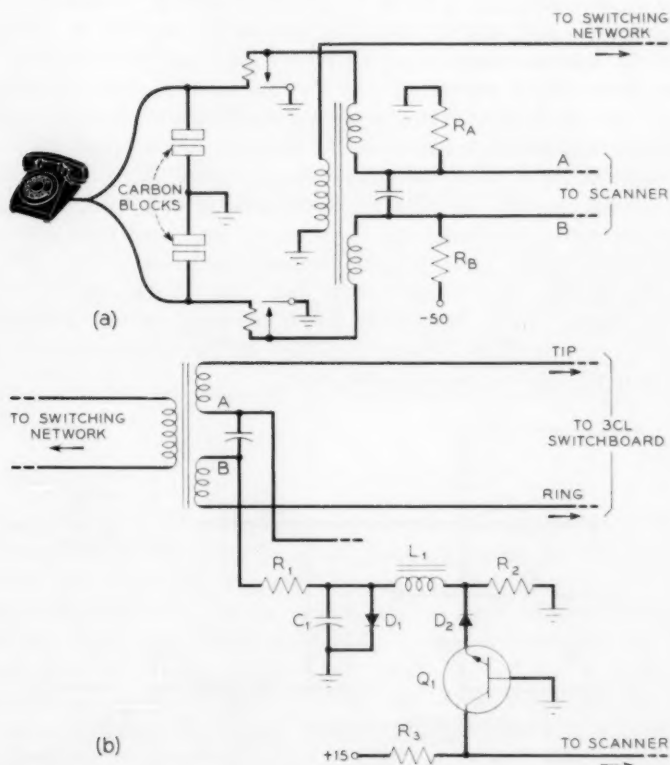
Fig. 2 — (a) Customer line circuit; (b) portion of trunk circuit which develops
scanner voltage.

If the diode is held in the back-biased state, an interrogation pulse coupled through capacitors to the diode will be blocked; if the diode is forward-biased, the pulse will be passed and can be detected. Although the actual "transmission gate" circuitry is somewhat more complicated and will be described in detail later, it is shown in its simplest form in Fig. 3.

1.2 Scanner Operation

The functional arrangement of the scanner is shown in Fig. 4. Transmission gates are located at the points of intersection of 32 horizontal wires and 32 vertical wires arranged in a square array. To determine which of the 1024 gates is to be interrogated, central control addresses the horizontal and vertical wires that intersect at the gate in question. A five-bit address selects a path through the input selector from the pulser to the driver associated with the desired horizontal. The output selector sets up the path from the output amplifier and lock-up amplifier associated with the vertical lead to the detector. A 1-microsecond 6-volt interrogation pulse called the enable scanner pulse (ENS) is then sent from central control. The ENS resets the lock-up amplifiers, erasing information stored in them from the previous scan, and then starts its journey through the input circuits, the addressed transmission gate (if properly biased), the output circuits and, from the detector, back to central control. Delay through the pulser and input selector guarantees that the resetting of the lock-up amplifiers will be completed before the pulse gets through the transmission gates.
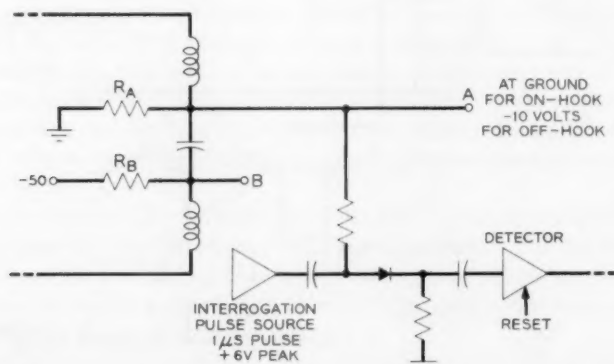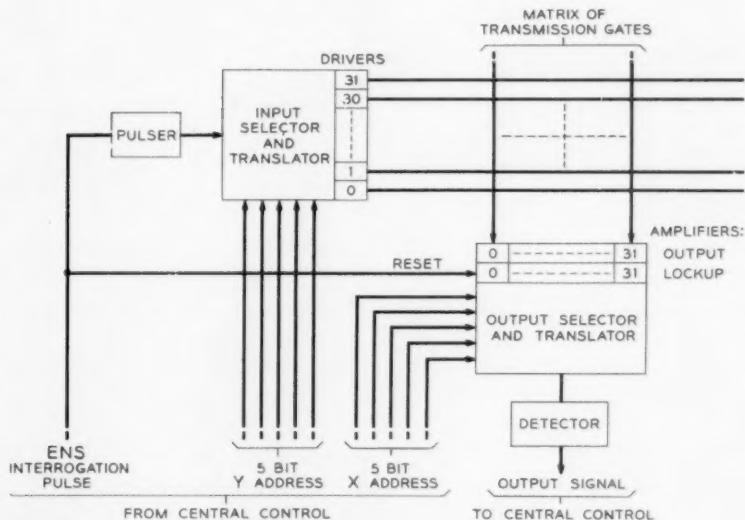

Fig. 3 — Simplified transmission gate.

Fig. 4 — Block diagram of the scanner.

It should be noted that any one horizontal lead is associated with 32 transmission gates. All 32 will be interrogated by one ENS pulse, and their lock-up amplifiers will be either set or not set as dictated by the state of the circuit being scanned. The five-bit address to the output selector can then be varied to transmit the state of as many of the 32 lock-up amplifiers as desired to central control without sending another ENS pulse.

1.3 Scanning Rates

Central control operates the scanner in two different modes to obtain the information it needs in the most efficient manner. As long as a customer is not using his phone, his line remains in a quiescent state; central control would waste program time if the line were scanned too often. Similarly, when a customer is talking, his line is in another dc state which does not change rapidly. Only when the customer is dialing do rapid changes in the dc state of his line occur. Thus, all dialing lines must be scanned much more rapidly than idle or busy lines. Fig. 5 shows the idealized line voltage at point A of Fig. 2(a) as the customer picks up his phone, dials a number and carries on a conversation. A supervisory scan samples his line (and all other lines, trunks and test points in the system)
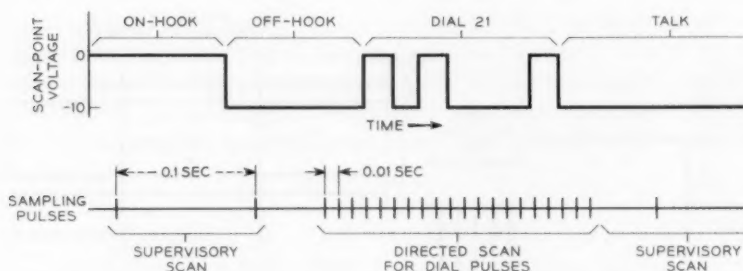
Fig. 5.—Idealized dial pulse voltage at scan point, as sampled by the scanner.

every 0.1 second. When an off-hook is discovered, the central control is programmed to sample the line 10 times faster (every 10 milliseconds) until all dial pulses have been registered. When the connection is set up, central control reverts to the less frequent supervisory scan which will detect hang-up when it occurs and signal the removal of the connection.

Fig. 6 shows the basic time cycle of central control, which repeats every 0.1 second. In each 5-millisecond interval central control obtains from the scanner a sample of the state of half the dialing lines in the office. Then $\frac{1}{20}$th of all lines, trunks and test points are scanned. During the remainder of the 5 milliseconds, central control performs other functions such as outpulsing, setting up connections through the network or test programs. The time devoted to sampling each line has to be very brief to permit central control to operate serially in this manner; the scanning interval approaches 2.5 microseconds.
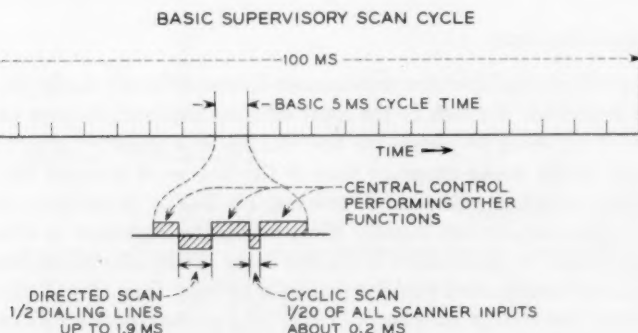
Fig. 6 — Time relations in scanner operation.

1.4 *Use of Scanner Information in Central Control*

Each line in the office is associated with, in addition to a particular scanner input, two binary cells in the temporary memory (barrier grid store[2]). These so-called line spots define the three basic states of a line as follows: (a) idle: 00; (b) talking: 10; (c) a register is assigned (line is dialing): 01.

When any given line is being scanned, central control is, at the same time, also examining its line spots. When the line is in the idle or talking state, it is scanned at the relatively infrequent supervisory rate, as described above. Should the output of the scanner show that a line is off-hook when its line spots are in the 00 state, the customer has taken his phone off hook since the last time the line was scanned. Similarly, if the line is on-hook but has 10 for its line spot state, central control knows that the customer has just hung up. In either case, central control will take the action demanded by the nature of the change and then bring the line spots up to date.

When off-hook is first detected, central control assigns the line an additional spot group in a section of the barrier grid store referred to as an originating register. In addition, the line spots are changed to 01. On the next supervisory scan, 01 in the line spots indicates to the central control that it is to ignore the scanner output regardless of what it is. The dialing line will be examined every 10 milliseconds through a specially directed scan where the line's address is obtained from the originating register. The over-all program involving the directed scan and the originating register counts, translates and stores dial pulses, and keeps the originating register up to date. It also times interdigital periods and, finally, signals the completion of dialing so that the connection can be established to the called party by means of another program.

II. PROBLEMS INVOLVED IN SCANNING

External conditions impose limitations on scanner design in two ways. First, the scanner must conform to over-all system requirements. Second, outside plant and customer subset circuits provide dial pulses and supervisory voltage levels that vary considerably from one loop to another. The scanner must be capable of recognizing and properly quantizing this wide variety of signals.

2.1 *Systems Considerations*

Of the systems factors involved, the most important is time. If the sampling action of the scanner is to provide central control with all the

signal information transmitted by the customer's dial, the sampling rate must be at least twice the highest signaling frequency. The subsets to be used will have 20-pps dials which, when dial speed and break-time variations are considered, will have a minimum make-time of 15 milliseconds. This established the 10-millisecond scan rate for dialing lines. The supervisory scan, although somewhat slower, must be fast enough to detect off-hook in time to permit central control to set up an originating register spot group in the barrier grid store before dial pulses are received. One scan every 100 milliseconds is satisfactory for this purpose. The maximum rate at which the scanner can be operated is determined by delays through blocks of equipment operated in tandem, RC discharge times and other circuit limitations.

In the supervisory scan, lines and trunks are sampled in a fixed order determined by adding 1 to the address of the last line or trunk inspected. In the dial pulse scan, however, consecutive addresses do not necessarily have any relation to each other. Thus, the scanner must provide random access to all inputs.

The scanner address and ENS leads must be compatible with the outputs from central control, and the scanner output signal must be capable of operating central control circuitry.

2.2 *Customer Circuit Considerations*

With regard to customer loop variations, Fig. 7 shows a worst tolerable condition from the dc standpoint. The subset itself has a maximum resistance of 1800 ohms in series with the dial and switch-hook contacts.
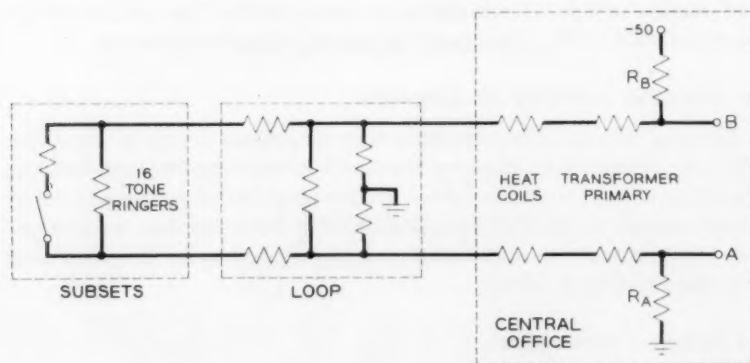

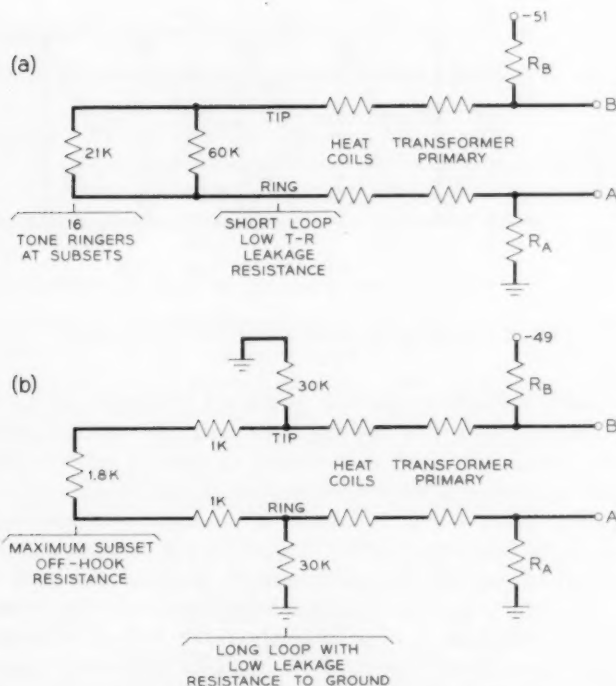Fig. 7 — Dc circuit for a long customer loop.

Fig. 8 — (a) Worst on-hook dc condition: interrogation pulse tends to be blocked when it should pass through gate; (b) worst off-hook dc condition: interrogation pulse tends to get through when it should be blocked.

Tone-ringers, which replace bells in the experimental systems' subsets, have about 340,000 ohms resistance.* An eight-party line with two subsets per customer makes the total shunt tone ringer resistance as small as 21,000 ohms. Because the maximum subset current must be limited to 13.5 milliamperes, outside plant leakage currents must be kept very small by comparison; the minimum leakage resistances permitted are 60,000 ohms from one conductor to the other and 30,000 ohms from either conductor to ground. These are the worst conditions which can be permitted if extreme values of component variation within the scanner occur and maximum longitudinal induction (15 volts, peak to peak) is present.

---

* While they are being operated, tone ringers have a much lower impedance. Thus the scanner must be programmed to scan ringing lines between bursts of ringing current to detect answer correctly.
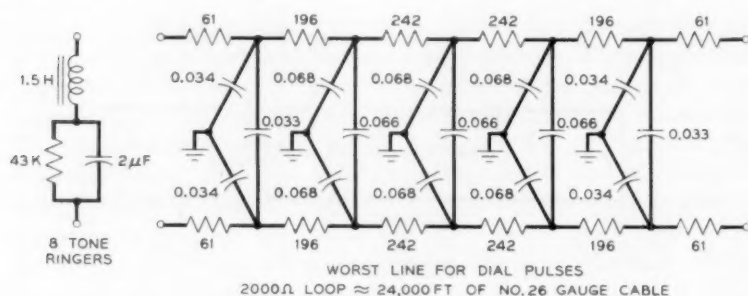
Fig. 9 — Circuits to simulate loop and tone ringers.

Supervisory voltages at the scan points are a function of the resistances discussed above. Two limiting cases exist: the condition in which the scanner comes just short of mistaking an on-hook condition for off-hook, and *vice versa*. Fig. 8 summarizes these dc loop conditions.

In addition to the circuit resistance, shunt capacity of the loop and reactance of the tone ringers affect dial pulses. Fig. 9 shows an artificial line designed to simulate the worst tolerable path for dial pulse transmission. The total conductor resistance is 2000 ohms and the distributed capacitance corresponds to approximately 24,000 feet of exchange area cable. Fig. 9 also shows an equivalent circuit for eight tone ringers. All these reactive components tend to round off the sharp corners of dial pulses and provide the scanner with dial pulse trains that, in the worst cases, look very much like sine waves.

Noise picked up by the loop can distort dial pulses badly. One of the worst noise problems is produced by 60-cycle voltage induced in open-wire telephone lines by power lines that run near them. Fairly large voltages from this source are often encountered, but fortunately they are, in most cases, of the same magnitude and phase in both the tip and ring conductors. Fig. 10(a) shows a train of dial pulses from a distant customer, in terms of the voltage across R_A of Fig. 2(a). Fig. 10(b) shows the dial pulses submerged by 60-cycle pick-up. If a differential oscilloscope is placed across points A and B in Fig. 2, however, the 60-cycle voltage will be balanced out. In addition, because of the direction of the battery current through the resistors, the dial pulse voltages add; this almost doubles the dial pulse signal. The scanner has been designed to use the voltage between points A and B to balance out 60-cycle interference and improve margins by taking advantage of the increased pulse height available.

III. SCANNER CIRCUITRY

If longitudinal induction were not a factor, the simple transmission gate of Fig. 3 could be adapted very simply to the practical problem of line scanning. Although not all loops are disturbed by longitudinal induction, it is desirable to make all line circuits identical and capable of eliminating the undesirable effects of longitudinals to permit straightforward administrative procedures.

3.1 *Trunk Gate*

Test points within the central office and scan points associated with most trunk circuits are not subject to induction. Fig. 11 shows the transmission gate developed for such circuits. It is relatively simple to divide the matrix into two groups of gates, one containing only balanced gates and the other, for trunks and special circuits, containing either the balanced gate or the one shown in Fig. 11.

The only differences between the circuit of Fig. 3 and that of Fig. 11
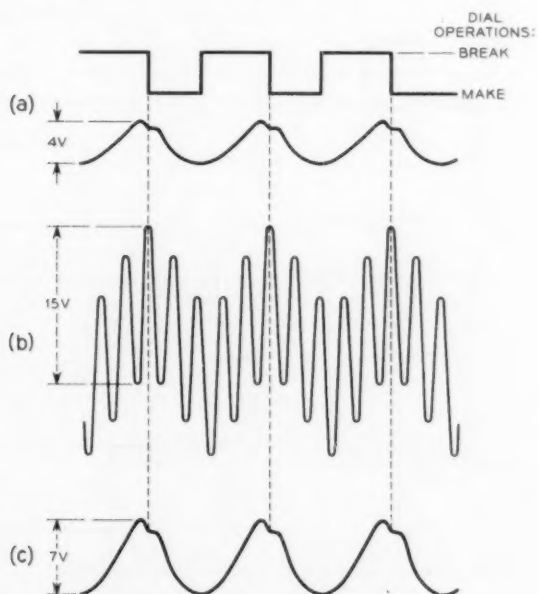
Fig. 10 — Scan point signals: (a) voltage at point A of Fig. 2, dial pulses from long loop; (b) same as above, with 60-cycle longitudinals present; (c) voltage at point A minus voltage at point B of Fig. 2, with 60-cycle longitudinals present.

are noise-filtering capacitors and a very small bias current which keeps the diode slightly forward-biased when point A is held near +15 volts. This bias aids in keeping charge from ENS pulses from being trapped on the cathode side of the diode and altering the diode bias.

3.2 *Balanced Line Gate*

The balanced line gate is shown in Fig. 12. Longitudinal voltages are coupled in phase to both ends of the diode leaving the net dc bias voltage unaffected. Capacitor C_1 blocks the central office battery but is a low impedance to both longitudinal voltages and dial pulses. Capacitors C_4 and C_5 are very high impedances to low-frequency signals, but transmit the 1-microsecond ENS pulse readily. They can be made to have large voltage ratings in small physical sizes and, coupled with hit filters R_1C_2 and R_3C_3, effectively protect other scanner circuits from lightning and other noise voltages which may be picked up by the loop.
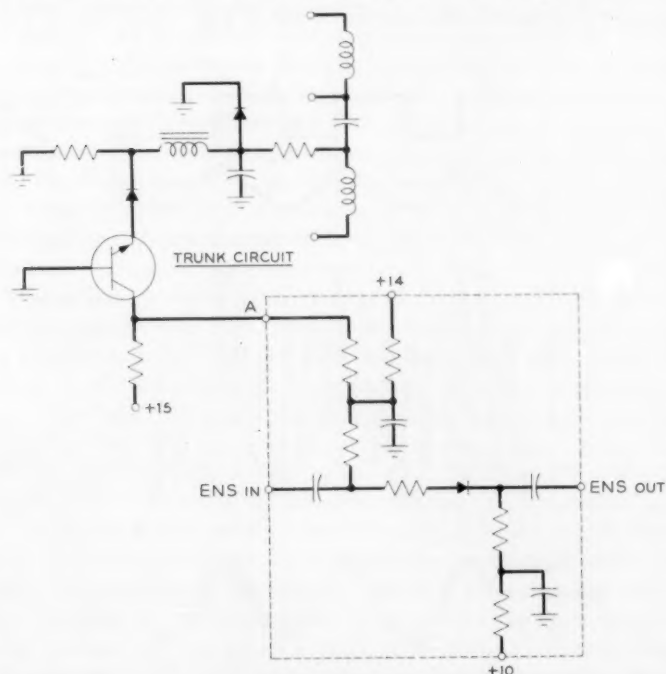

Fig. 11 — Transmission gate for trunk circuits and test points.

Because 32 gates associated with a horizontal matrix lead are pulsed in parallel, R_7 is inserted to equalize variations in diode forward resistance; R_4 is the dc return path for the diode, and offers a convenient place to forward bias the diode to reduce the effects of trapped charge; R_2 is required to maintain ac circuit balance.

The balanced line gate has been exhaustively tested for effects of component variations, loop variations, longitudinal induction and lightning hits.

3.3 Input Selector

Although the matrix of line and trunk gates is the heart of the scanner, the access circuitry which enables central control to sample one particular scan point out of the 1024 available requires careful consideration. The ENS pulse must be steered to one of 32 horizontal matrix leads, and a path from one of 32 vertical matrix leads must be established to the detector which transmits the quantized output signal back to central control.

The input selector is a transistor "tree" as shown in Fig. 13. In the first stage, only one of the eight transistors has its base grounded; the input translator holds the other seven base leads at a positive voltage. A positive pulse applied to the eight emitters in parallel thus finds only
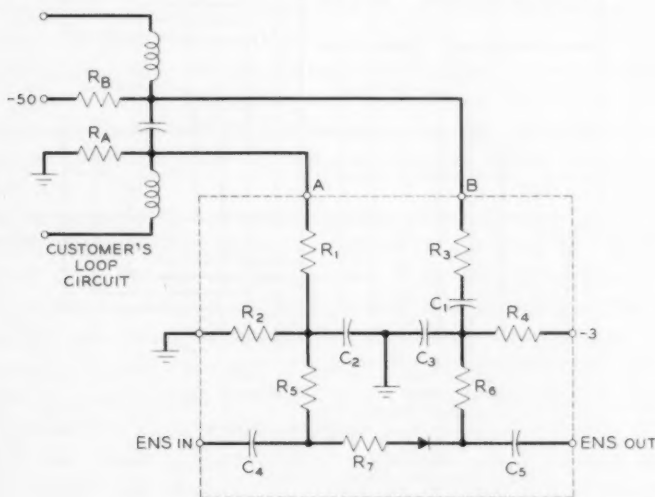

Fig. 12 — Balanced transmission gate for customer loops.

one emitter-base junction forward-biased. In the second stage of the transistor tree, the positive pulse finds only one of four paths open to it as a result of translator base voltages, and the one-out-of-32 selection is performed.

The input translator, part of which is shown in Fig. 14, is made up of diodes and resistors. It is driven, through several sets of gates and amplifiers, by address flip-flops in central control.

Once the path to a matrix horizontal lead has been established through the input selector, the ENS pulse, reshaped and amplified by suitable circuitry, can be applied.

3.4 *Pulser and Drivers*

Reshaping is performed by the pulser, shown at the left in Fig. 15. The first stage is a blocking oscillator and the second is a transistor
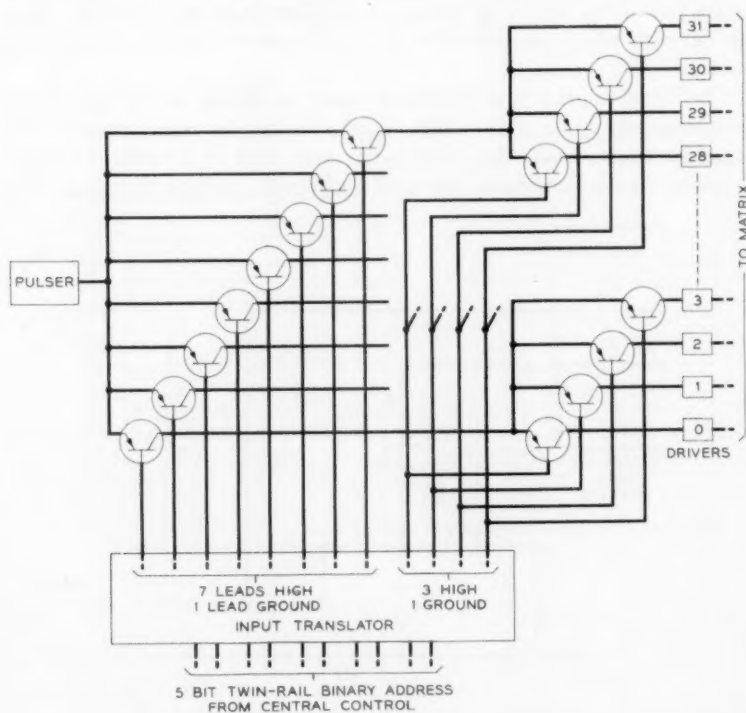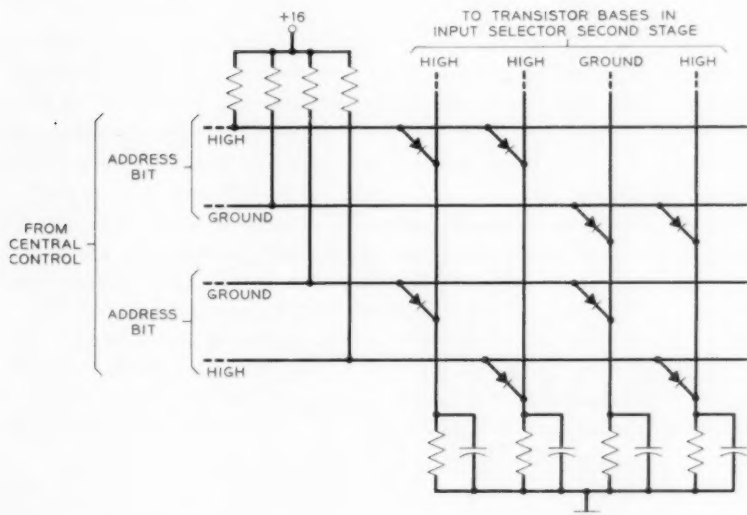

Fig. 13 — Input selector.

Fig. 14 — Portion of input selector translator.

switch. When Q_1 is triggered by an ENS pulse from central control, regeneration through the pulse transformer T_1 drives Q_1 into saturation; Q_2, normally biased to cutoff by the forward voltage drop across D_2, is driven to saturation by the current pulse in the third winding of the blocking oscillator pulse transformer. The collector voltage of Q_2 has a rise time on the order of 0.2 microsecond; this fast-rising positive pulse is applied to the input selector and steered to the desired driver, shown at the right of Fig. 15.

The driver is another transistor switch. The high-current pulse from Q_2 is stepped up through pulse transformer T_2, saturates Q_3 and applies a 6-volt pulse through zero impedance to the matrix horizontal. The 6-volt supply must be accurately regulated; if it drifts higher, the ENS pulse may pass through off-hook line gates that are back-biased by the minimum permissible amount. If it drifts lower, the ENS pulse may not be able to pass through line gates where leakage currents have raised the on-hook bias to the maximum level.

Both the pulser and driver have capacitors on all voltage supply leads to filter noise and eliminate ringing. Protective resistors are inserted wherever required in transformer leads to protect the transformers against possible transistor short circuits.
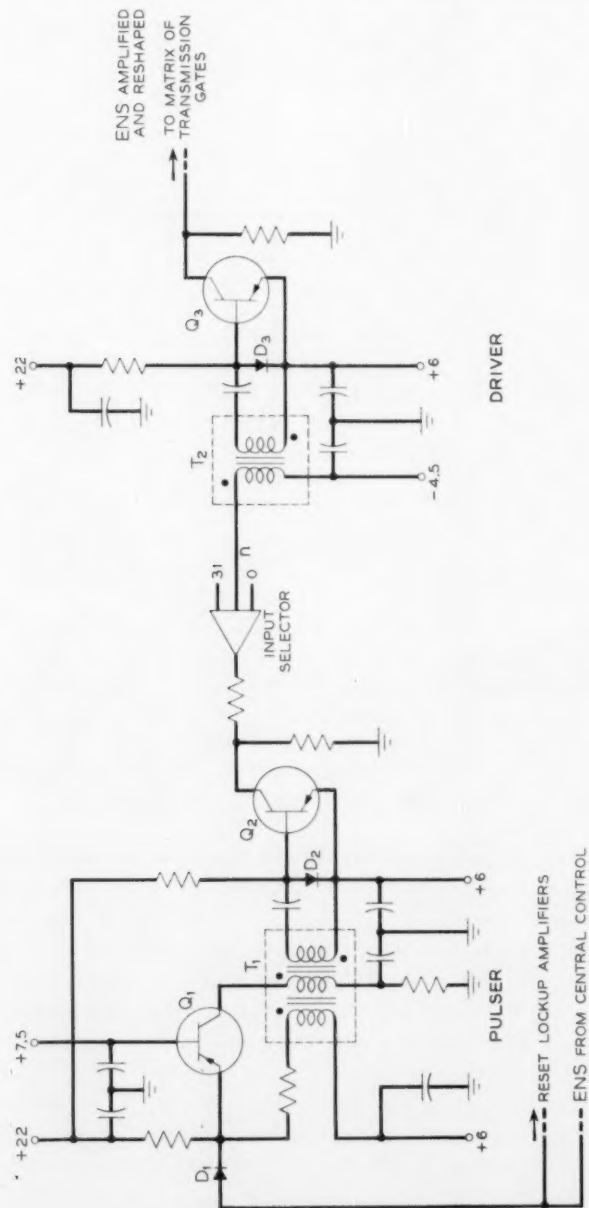
Fig. 15 — Pulser and driver.

3.5 *Output and Lockup Amplifiers; Reset Circuits*

The driver can supply 175 milliamperes to a horizontal lead of the scanner matrix. This current divides among the 32 line gates associated with the horizontal and, due to shunting elements within the matrix and the gate circuitry, output current on a vertical lead may be as small as 1 milliampere, with a duration of 0.5 microsecond.

This output current must be amplified before it can be used effectively. In addition, some threshold value must be established to block noise. These requirements are met by the output amplifier shown in Fig. 16, one of which is associated with each matrix vertical.

Thirty-two transmission gate outputs connect to each matrix vertical; only one of these gates is pulsed at a time. To minimize the shunting effects of the 31 unpulsed gates, the first stage of the output amplifier is a grounded-base transistor with the typical low impedance of that configuration. Acting as a linear amplifier, the first stage transmits current
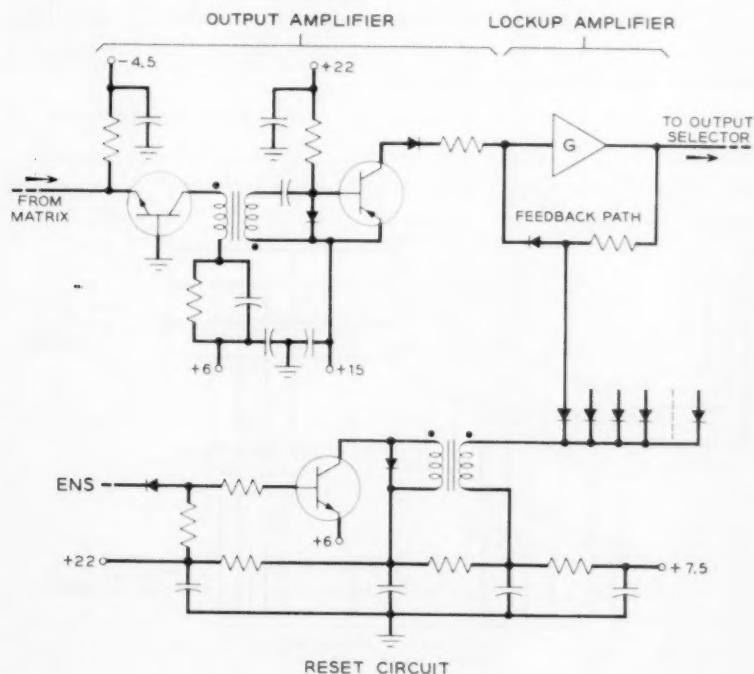

Fig. 16 — Output amplifier, lockup amplifier and reset circuit.

to the second through a pulse transformer. The second stage is the now familiar scanner transistor switch, biased off by the forward drop across the diode. Noise discrimination takes place here; if the input current pulse is less than 1 milliampere, the output pulse from the transformer is unable to turn on the second transistor. For a large enough input, however, the second stage operates and can easily switch the lockup amplifier which follows.

The lockup amplifier, also shown in Fig. 16, is a standard gate amplifier[3] with external feedback added. In the reset position, its input is clamped at +6 volts. A pulse from the output amplifier drives the lockup amplifier input to +7.5 volts, where it is clamped by a diode and maintained by the feedback loop after removal of the pulse.

Each time central control sends an ENS pulse to the scanner, all lockup amplifiers are reset. The ENS pulse triggers a transistor which applies a square negative pulse through a pulse transformer and isolation diodes to the feedback loops of the lockup amplifiers. The transformer is chosen to give a pulse long enough to guarantee resetting the lockup amplifiers, but short enough to be removed by the time the signal is available from the output amplifier.
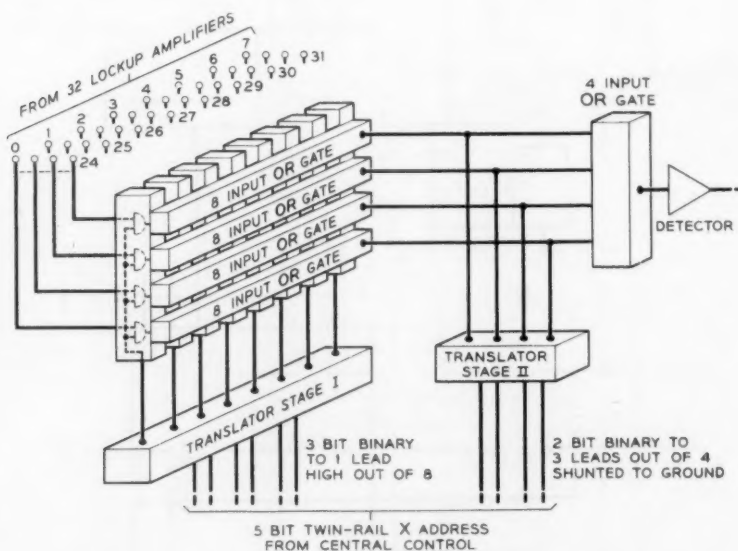

Fig. 17 — Output selector.

3.6 *Output Selector and Detector*

The output selector shown in Fig. 17 is a small diode tree scanner designed to connect lockup amplifier outputs through to the detector under the direction of central control. Three bits of the five-bit address choose one block of four output amplifiers. The other two bits select which of the four is to be observed by shunting the other three leads to ground.

The detector, shown in Fig. 18, provides a twin-rail output to central control. When the addressed lockup amplifier is set, Q_1 conducts and Q_2 is cut off; base-emitter voltages are developed across diodes D_1 and D_2, respectively, the anode of D_1 being slightly above ground and the cathode of D_2 slightly below ground. When the lockup amplifier is not set, it clamps the output selector output to ground. The voltage divider at the base of Q_1 holds Q_1 cut off, and Q_2, its base clamped just below ground and its emitter connected to -4.5 volts through a small resistor, conducts; Q_3 and Q_4 are emitter-follower amplifiers which drive the cables back to central control. The detector has a very fast rise time and is capable of developing enough power to charge up the distributed capacity of the output cables with a minimum of delay.

3.7 *Construction Details*

Scanner circuits are built on printed wire boards and several circuit packages are shown in Fig. 19. The balanced line-gate card, containing
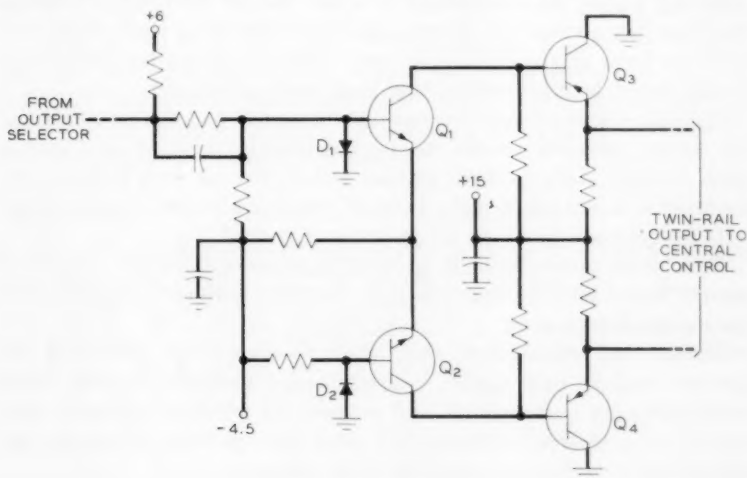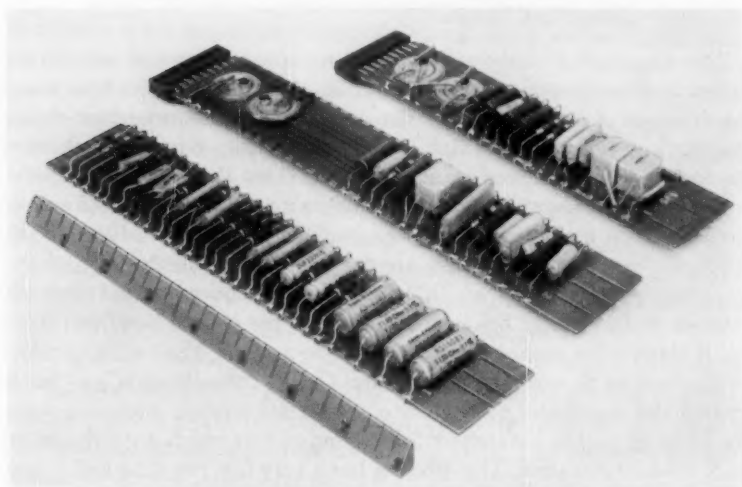

Fig. 18 — Detector.

Fig. 19 — Line gate, output amplifier and driver packages (left to right).

two gate circuits and two pairs of line resistors, is shown in the foreground.

Printed wire boards are plugged into jacks arranged on horizontal mounting plates. All connections between circuits are made by wiring from one jack to another. Shunt capacity to ground in such wiring can be fairly large and special pains were taken to design low-impedance circuits where long interconnecting leads were anticipated.

The high current pulses required by the matrix and the sensitivity of the output amplifiers make special procedures necessary to insure a quiet ground. Heavy copper ground straps at each jack position are mounted in every other space between mounting plates, giving all circuits ready access to a very low impedance ground.

With these precautions, it is felt that scanner reliability has been greatly increased and that coupling between individual circuits will cause negligible trouble.

The line and trunk gates (which make up most of the volume of the scanner) contain only diodes, resistors and capacitors. Because these components are quite reliable and because the failure of any one component in a gate circuit will not affect more than one input to the scanner, duplication for reliability purposes is not necessary.

On the other hand, each piece of circuitry used to obtain access to the matrix gates is common to many lines and must be duplicated. Two

complete blocks of access circuitry are provided and, in the event of trouble, all the circuits (pulser, input selector, drivers, output amplifiers, etc.) associated with a given matrix are automatically switched out at once and replaced by similar standby circuits. Wire-spring relays under the control of the administration center effect this transfer.

A 1000-line scanner, complete with duplicated access circuitry, can be built conveniently in one electronic switching system equipment cabinet (2 x 2 x 7 feet), with enough space left over to hold the electronic power supply specially developed to provide talking current for customer lines.

IV. CONCLUSIONS

The scanner described in this paper operates serially to sample the voltage levels at 1024 input terminals. These levels are quantized to 1 or 0 and transmitted to central control at a fast enough rate to insure proper operation of the telephone system. Most scanner inputs are connected to customer loops. If loop variations are maintained within the specified limits, the scanner will function properly. Other inputs are used to scan trunk circuits and diagnostic test points within the system.

As part of the rather complete laboratory model of the experimental electronic switching system, a skeletonized scanner is working satisfactorily.

V. ACKNOWLEDGMENTS

REFERENCES

1. Joel, A. E., An Experimental Switching System Using New Electronic Techniques, B.S.T.J., **37**, September 1958, pp. 1091–1124.
2. Greenwood, T. S. and Staehler, R. E., A High-Speed Barrier Grid Store, B.S.T.J., **37**, September 1958, pp. 1195–1220.
3. Yokelson, B. J., Cagle, W. B. and Underwood, M. D., Semiconductor Circuit Design Philosophy for the Central Control of an Electronic Switching System. B.S.T.J., **37**, September 1958, pp. 1125–1160.

A Signal Distributor for Electronic Switching Systems

By L. FREIMANIS

(Manuscript received July 7, 1958)

The purpose, design and operation of the signal distributor are described in this paper. The signal distributor is the unit in the experimental electronic switching system which is used to distribute action signals. A semiconductor gate selector, driving a large group of memory flip-flops is used to generate the necessary action signals from a high-speed binary address. Factors governing the design are outlined and some alternate solutions compared.

I. INTRODUCTION

The signal distributor is the means of distributing action signals in the experimental electronic switching system (Fig. 1). The central control processes the necessary information, employing high-speed low-power electronic circuits. The signal distributor converts the high-speed low-power signals received from the central control into low-speed high-power signals required for certain operations in the system. The control of supervision or dial pulsing in trunk circuits constitutes the major part of these operations, and control of various maintenance circuits within the administration center is one of the other miscellaneous operations in the electronic switching system that is performed by the signal distributor.

The signal distributor consists of a selector driving a large group of flip-flops with associated amplifiers (Fig. 2). The amplifiers have sufficient power capacity to operate relays where necessary. The flip-flops permit output signals of long duration. They act as memory elements, accepting two high-speed signals from the selector which designate the starting and ending times of the output signal. The selector performs the translation of a binary address received from the central control.

II. FUNCTIONAL REQUIREMENTS

The binary address is presented to the signal distributor as a particular combination of potentials on 21 address leads (Fig. 2). Twenty of
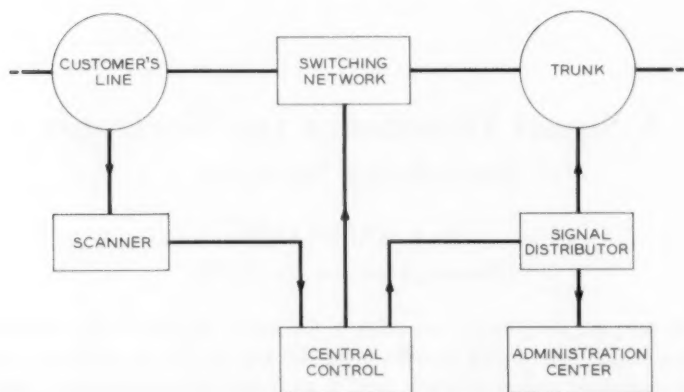
Fig. 1 — Simplified block diagram of an experimental electronic switching system.

these leads are grouped in ten pairs, the 21st is the enable lead. The potentials on any of the ten pairs are always conjugate: if one of the leads in a pair is at ground potential or in the passive state, the other will be at a positive potential or in the active state. Consequently, each pair can assume two distinctly different states and, for the ten pairs of address leads, there are $2^{10} = 1024$ possible combinations of potentials. A combination is assigned to every order. The order must be carried out by the signal distributor whenever the particular combination appears on the ten pairs of address leads and the enable lead becomes active.
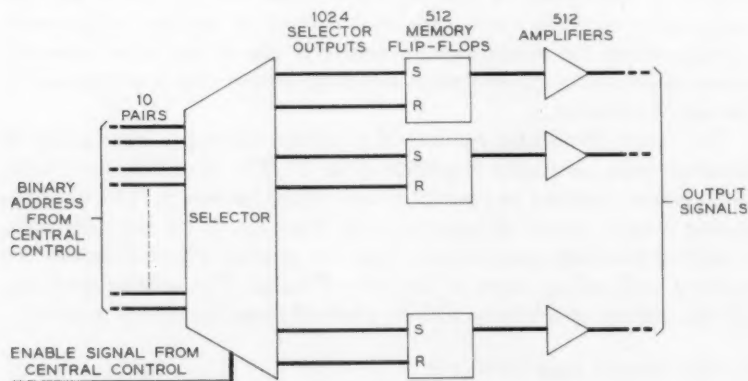

Fig. 2 — Functional diagram of the signal distributor.

For the control of each slow-speed output signal path two orders are necessary, one to start a particular action and another to terminate it. Therefore, the 1024 available combinations or addresses are sufficient to control 512 different output signal paths. After receiving the binary address, the selector must produce an active signal, suitable for activating a memory flip-flop, on one out of 1024 leads. This is a process of combinational logic and is commonly known as binary to one-out-of-n translation.

Various devices and circuits can be employed to accomplish the translation. A particular design, however, must be compatible with the power levels available at the inputs and required at the outputs. It also must be able to respond to the fastest expected address repetition rates. Furthermore, the most economical design that still meets the basic requirements and is sufficiently reliable should be attempted.

The input or address signals to the selector are presented by transistor amplifiers in the central control. Antisaturating techniques are used in these amplifiers, which limit their current carrying capacity to 30 milliamperes. Also, because of antisaturating circuits, the amplifier outputs never return to an exact ground potential in the passive state. The potential in the active state is limited by the breakdown voltage of the transistors used. An active potential of 16 volts is considered a safe value.

The fastest expected address repetition rate is 7.5 microseconds. This means that the potentials on at least one pair of address leads may be reversed every 7.5 microseconds. Under certain conditions the potentials on all ten pairs may be reversed at this rate. The selector must be able to accept addresses and produce an active output on one of its 1024 output leads every 7.5 microseconds. Furthermore, the memory flip-flops should be able to change their states within this time.

The memory flip-flops in the signal distributor are of the same design as those used in the logic circuitry of the central control. They require a 5-milliampere current pulse for less than a microsecond on their input leads for setting or resetting. The memory flip-flops themselves have limited output-power capacity. A single-transistor amplifier is provided at the output of each flip-flop. The transistor in this amplifier is allowed to saturate, thus permitting comparatively large currents to be switched. In fact, with proper protection against inductive surge voltages, relays can be operated directly from the amplifier output. The amplifier also serves as a buffer between the output load and the memory flip-flop. The output loads may vary widely, depending upon the nature of the connected circuitry.

The outputs of the signal distributor have to operate various types of relays as well as different slow-speed electronic circuits. For the type of memory flip-flop used in the signal distributor, it is essential that there always be a constant load on the flip-flop outputs. If this load is permitted to vary, the operation of the flip-flop will be seriously affected. The normal driving pulse may not be sufficient to set or reset the flip-flop, or it may be impossible to change the state of the flip-flop at all. The amplifier presents a constant load to the flip-flop, independent of the load variations to the amplifier. In this way, the requirements on the output signal of the selector are the same for all 1024 selector outputs. The output signal of the selector must be sufficient to change the state of the memory flip-flop. A current pulse of 5 milliamperes amplitude is required, and it must overcome a threshold of about 5 volts, which threshold is built into the flip-flop to guard against triggering from noise. The pulse must exist at the input terminal of the flip-flop for at least 0.3 microsecond to assure positive triggering.

III. DESIGN CONSIDERATIONS OF THE SELECTOR

Because of the input and output power levels, as well as the required address repetition rates, semiconductor gates are best suited for performing the function of binary to one-out-of-n translation in the signal distributor. The required repetition rate of 7.5 microseconds is sufficiently low to permit the use of saturating transistor circuits wherever amplification is necessary. Saturating circuits can switch larger currents than can circuits using antisaturating techniques, and they are simpler and more efficient. The gates can employ either transistors or diodes, and can be arranged in a single or multistage pattern.

In deciding whether transistors or diodes should be employed as gates, and how many stages of logic should be used, certain contradicting factors must be considered and a compromise solution found. In general, the number of devices should be minimized. In doing this, the relative cost and reliability of diodes versus transistors must be taken into account. Multistage logic reduces the number of gates, but it may require interstage amplifiers to compensate for loss of level. The total dc power consumption is an important factor; in large translators this can be greatly reduced by using resistors as gate elements and applying voltages only on the gates that are active at a given time. For a practical design, particularly one that is to be used in a telephone office, it is highly desirable to keep the number of different circuit units to a minimum. As the number of stages is increased, more different

circuit units are required. This complicates maintenance problems, particularly if replaceable plug-in type circuit units are to be used. A single-stage translator using diodes as gate elements and transistors as drivers or amplifiers has the advantage of simplicity. Only two different circuit units are required, because all the gates are the same, and there is no need for interstage amplification. For translators of small size, single-stage logic is the preferred solution.

There are definite practical limitations to single-stage diode translators. These limitations become increasingly more pronounced as the size of the translator increases. The number of diodes required in a single-stage diode translator is $n2^n$, where n is the number of binary digits. For a ten-bit single-stage translator, 10,240 diodes are needed. Another limiting factor is the input-to-output current ratio or fanout. In order to provide a certain current at the active output, each driving circuit at the input must be able to switch $(2^n - 1)/n$ times that current. At the inputs of a ten-bit translator, a current 102 times larger than the active output current must be switched. This would require considerable amplification of the available input signals. The dc power that must be supplied to a single-stage diode translator is 2^n times the power required for one active output.

For translators of four binary bits and larger the number of diodes can be reduced by various arrangements of multistage logic. The fanout and dc power can be reduced by using either transistor or resistor-diode gates in place of straight diode gates. The selector of the signal distributor employs diodes and resistors as gate elements in a two-stage logic, with transistors being used in amplifier circuits to provide gain where necessary. Compared to a single-stage logic diode translator, the circuit is considerably more economical. The number of diodes is reduced from 10,240 to 1,344, the required dc power is 32 times smaller and the input fanout is decreased from 102.3 to 6.2.

Small additional savings in the number of diodes can be obtained by introducing a three-stage logic. This, however, would require a few more different circuit units. Not only would the total number of units be larger, the necessity of keeping spares for all types of circuit units used in a device would actually increase the number of diodes that have to be provided.

Within the framework of the outlined specific requirements, and considering the limitations of available devices, the design of the selector is an optimum with respect to the number and cost of devices, as well as to power consumption. Reliability can only be estimated when more data is available on the relative failure rates of transistors and diodes.

IV. DESIGN OF THE SELECTOR

The two-stage selector of the signal distributor (Fig. 3) may be considered as consisting of the following basic building blocks: (a) primary translators, (b) enable circuit, and (c) secondary translator, or 32 x 32 matrix.

4.1 *Primary Translators*

The first stage of translation has two identical five-bit diode translators. Five pairs of input leads, or half of the address, are connected to each primary translator. Each translator selects and activates one out of 32 output leads corresponding to the binary input. The translators
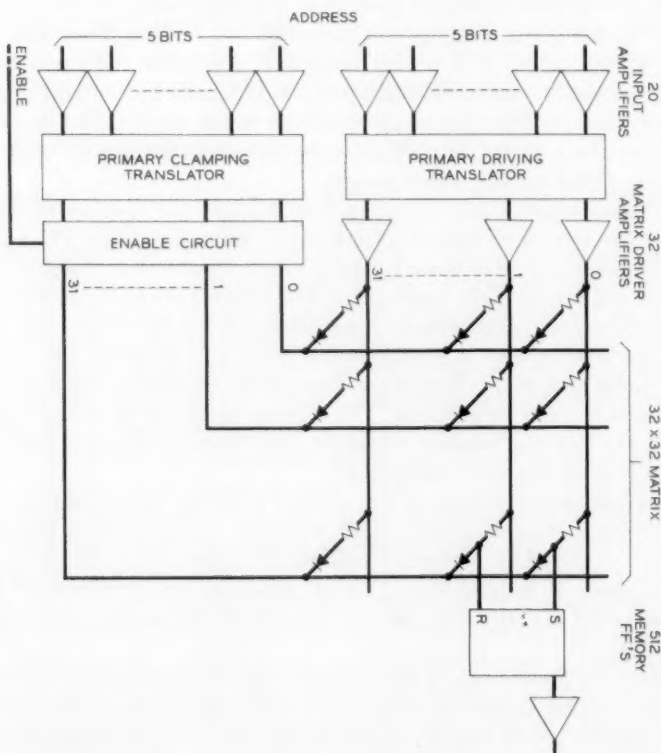

Fig. 3 — Block diagram of the signal distributor.

themselves consist of 160 diodes, five diodes for every output. The diodes are connected in an AND gate logic pattern.

At each input of the primary translators, a transistor amplifier or switch is provided. These amplifiers provide a low-impedance discharge path for the connecting cable capacitance. In this way, a regular switch-board cable can be used to transmit the fast-changing address signals from the central control to the signal distributor. The amplifiers also restore the voltage level of the address signals and provide a current sink at the inputs of the primary translators. The transistors in these amplifiers are allowed to saturate, thus introducing about 1.5 micro-seconds delay during the transition from the passive to the active state. This delay is mainly due to the storage time. The delay from the active to the passive state is only about 0.5 microsecond; this difference in transfer times is very desirable for the proper operation of the primary translators. It means that, whenever a pair of input leads is reversing its state, there is a time of about 1 microsecond when both leads are passive. This prevents any undesired outputs from the translator while the input address is changing.

The outputs of the two primary translators are used to operate the second-stage translator or the 32 x 32 matrix. Because of the nonsym-metrical two-input resistor-diode AND gate used in the matrix, the outputs of the two primary translators have to perform different func-tions. One of the primary translators has 32 amplifiers connected to its outputs, with each amplifier at an active output applying a 22-volt source to one of the inputs of the 32 x 32 matrix. The amplifiers are called matrix drive amplifiers, and the primary translator is designated the driving translator.

The other primary translator applies a ground clamp to 32 inputs of the matrix. Whenever an output of this translator becomes active, the ground clamp is removed from the corresponding matrix lead. This translator is called the clamping translator. All the 32 outputs of the clamping translator are normally kept at ground level by the enable circuit. Only after the enable circuit removes its ground from the 32 output leads can the active output remove the ground clamp from the selected matrix input.

4.2 *Enable Circuit*

The enable circuit prevents any active signals from the clamping translator reaching the matrix. A certain time after the enable circuit receives an active enabling signal from the central control, it allows the

signal to pass from the clamping translator to the matrix. The main purpose of this delay is to avoid false operations. When the address to the primary translators is changed, the selected output of the clamping translator becomes active before the previously active output of the driving translator disappears. The reason for this is the time delay in the matrix drive amplifiers. These amplifiers have to supply a comparatively heavy current to the matrix so that two stages of saturating transistors must be used. The time delay relationship of the outputs of the two primary translators is indicated in Fig. 4. If the active output from the clamping translator were permitted to pass to the matrix immediately, an undesired gate in the matrix would be activated. The enable circuit releases the clamping translator outputs 5 microseconds after the change of the input address, thus allowing sufficient time for the previously active matrix drive amplifier to reach the passive state. The clamping translator is released for 2 microseconds, which is sufficient time to change the state of the memory flip-flop.

The enable circuit must generate a delayed and stretched signal upon receiving the enable pulse from the central control. The circuit in Fig. 5
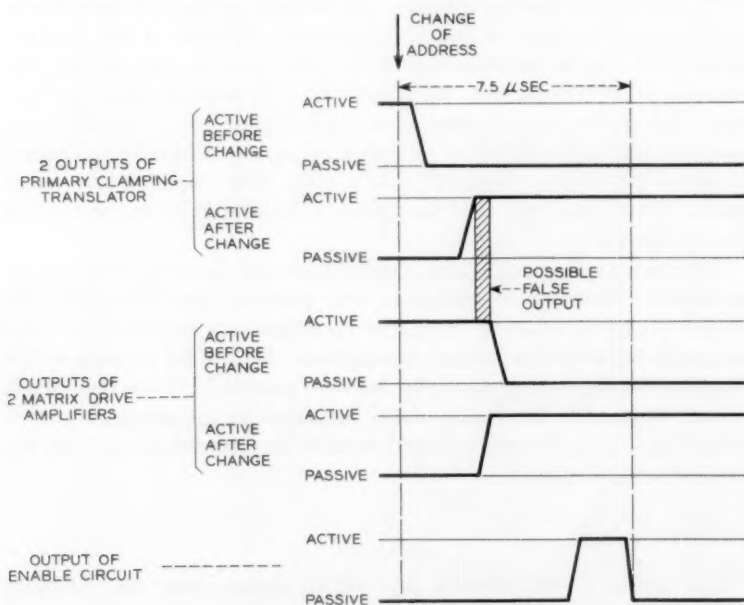

Fig. 4 — Time relations.

performs this function. The enable pulse arrives from the central control on a coaxial cable and, at the same time, the address is changed on the ten pairs of input leads. The cable must be properly terminated, because the enable pulse is only 0.3 microseconds long. Two self-resetting flip-flops are used to provide the timing; the first is set by the enable pulse and resets itself after 5.0 microseconds, with the trailing edge of its output setting the second flip-flop. This flip-flop resets after two microseconds; its output (after proper amplification) is used to enable the primary clamping translator. All 32 outputs of the clamping translator are provided with an extra diode each. The diodes are normally grounded by the amplifier of the delay circuit. When the delayed 2-microsecond signal arrives, the ground is removed and the active signal can pass from the clamping translator to the matrix.
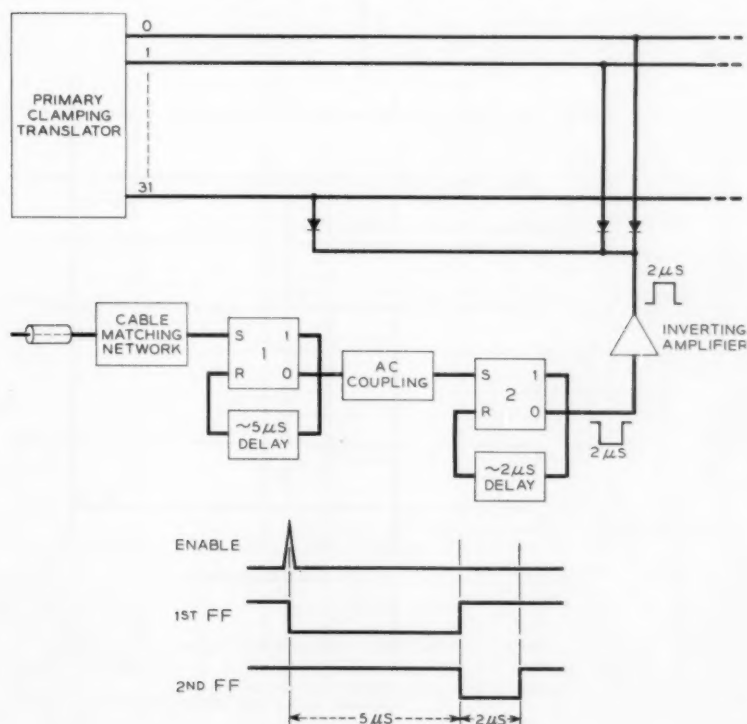
Fig. 5 — Enable circuit.

4.3 *Secondary Translator or 32 x 32 Matrix*

The second-stage translator may be considered as a 32 x 32 matrix consisting of 1024 resistor-diode AND gates. Simultaneous application of potential on the resistor and removal of ground from the diode is necessary to produce an output from the gate. An active matrix-driver amplifier applies 22-volt potential to a group of 32 resistors and the active output of the clamping translator removes ground from 32 diodes. There is only one gate that is common to both groups; in this, the current will flow from the matrix driver amplifier through the gate resistor into the memory flip-flop and change its state.
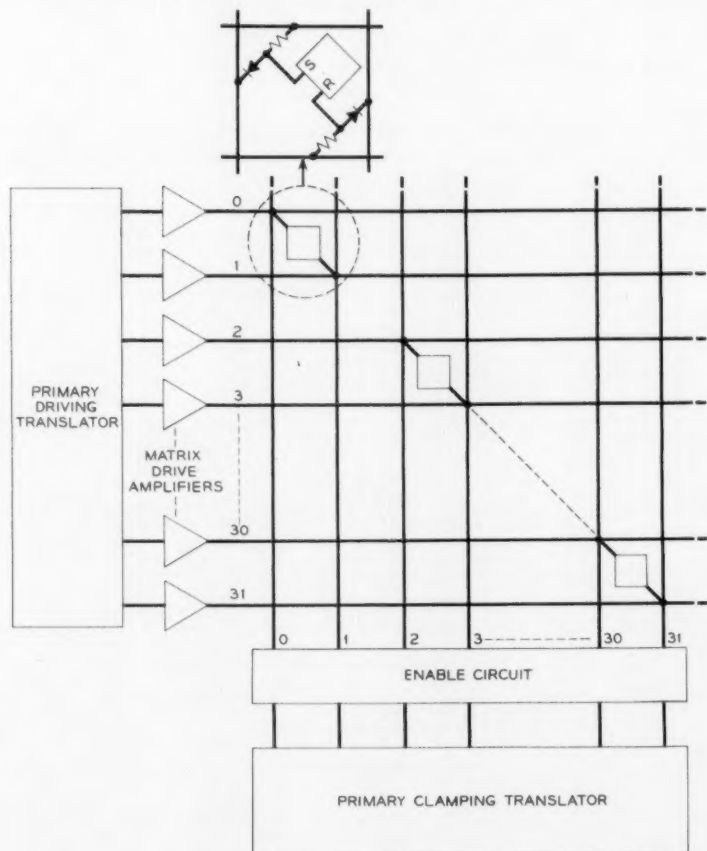
Fig. 6 — Test flip-flops.

V. TESTING AND MAINTENANCE

Every design of telephone switching equipment must provide sufficient reliability; a situation where failure of a component would disable the whole system cannot be tolerated. A total failure of the signal distributor would seriously affect the system. Not only would signaling on trunks be impossible, but many important testing and maintenance functions could not be carried out.

To insure reliability of the signal distributor, certain critical circuits are provided in duplicate and connected by relay contacts in such a way that one can be interchanged for the other in case of failure. Also, preventive maintenance can be performed in the standby circuit, greatly reducing the chance of failure. Only circuits in which a failure would disable more than one output are duplicated. Such circuits are the primary translators with the input amplifiers, the enable circuit and the matrix-drive amplifiers. Any failure within these circuits would disable at least
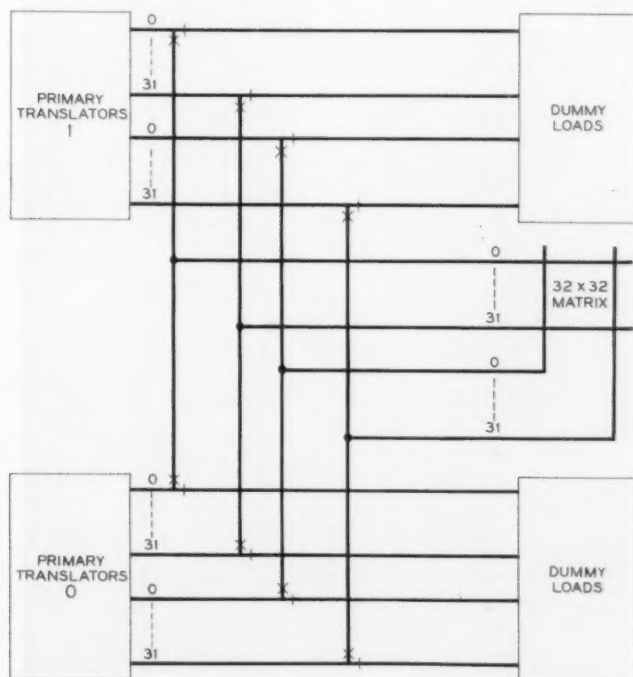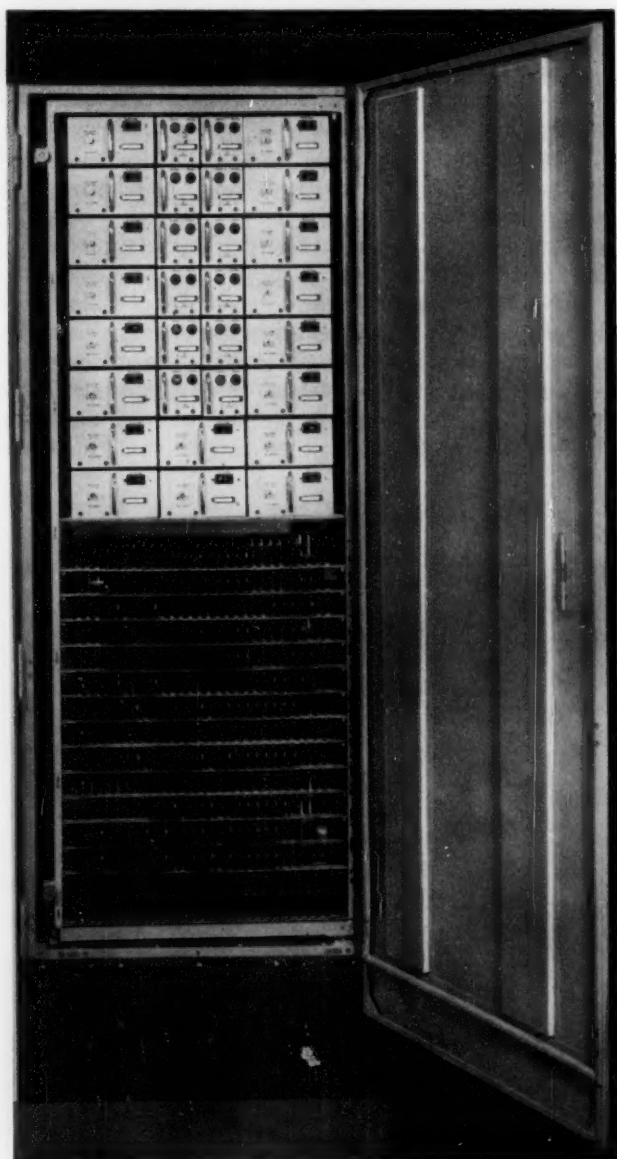

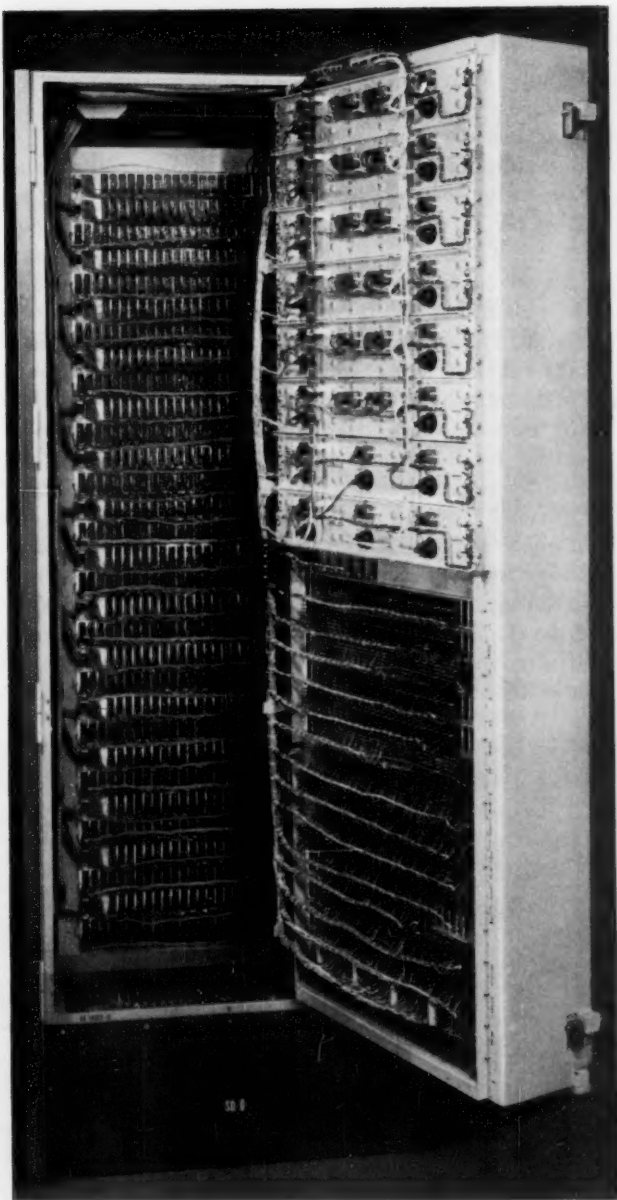Fig. 7 — Transfer circuit.

Fig. 8 — Signal distributor cabinet.

Fig. 9 — Cabinet with gate open

1417

32 output points. A single failure within the 32 x 32 matrix or a memory flip-flop, however, would affect only one output point. This is not considered serious and these circuits are not duplicated.

Before a standby circuit can be switched into operation, a failure of the active circuit must be recognized. For this purpose, a test routine is performed. At certain intervals the central control sends a sequence of addresses to the signal distributor to test its proper operation. Sixteen test flip-flops are connected at the outputs of both primary translators in such a way that they will detect an improper response of any of the outputs (Fig. 6). The central control observes the outputs of the test flip-flops by means of the scanner. Immediately after a failure is recognized, the faulty units are switched out of service and replaced by standby units. The standby units are also tested at regular intervals to assure that they are in operating condition. The actual switching of duplicated units is accomplished by relay contacts (Fig. 7). Interlock and sequencing circuits are provided, assuring continuity of clamping voltages. The same relays also apply dummy loads to the standby units in order to test them under loaded conditions. To prevent false outputs, no addresses are sent to the signal distributor during the switching operation.

Each of the duplicated units has its own power supply. Failure of a supply will be recognized as a failure of the whole circuit and the standby unit will be switched in. The rectifiers for the duplicated circuits are fed
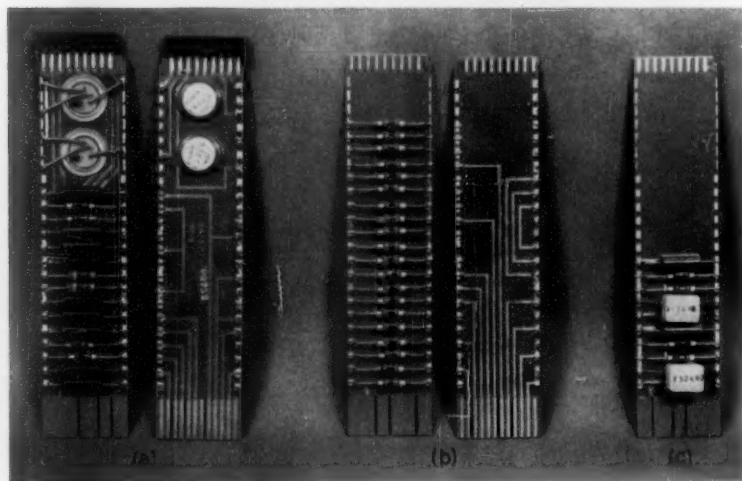

Fig. 10 — Typical circuit units.

from independent ac supplies, assuring continuity in case of an ac power failure. The continuity of power in the nonduplicated part of the circuit is assured by parallelling rectifiers.

VI. EQUIPMENT

The signal distributor is a self-contained unit. All the electronic circuits, power supplies and transfer relays are mounted in one enclosed cabinet. Fig. 8 shows the cabinet with front door open. The individual power supplies, duplicated primary translators and transfer relays are mounted on a swinging gate. Opening the gate (Fig. 9) provides for easy access to the terminals and wiring. The back of the cabinet contains all the output flip-flops with associated amplifiers.

The electronic circuits are mounted as units on replaceable printed wiring boards. Fig. 10 shows three typical circuit units: (a) is the transistor flip-flop used as the output memory element, with a shorting shoe provided to facilitate testing of the transistors outside of the circuit; (b) contains only diodes and constitutes part of the primary translator, the whole five-bit translator consisting of eight similar units; (c) is the coaxial cable terminating network.

The circuit units are slot-coded and plug into appropriate jacks. They can be easily replaced without interrupting operation of the unit. A centralized repair service of defective units is anticipated. This, together with the automatic trouble location, should greatly simplify the maintenance of the electronic switching system.

VII. ACKNOWLEDGMENTS

The equipment concepts which are the basis for the signal distributor design as well as the common control and scanner are the result of many contributions, notably those of H. A. Miloche and H. J. Wirth, Jr. The specific equipment design of the signal distributor was done by A. A. Nehez.

Application of Breakdown Devices to Large Multistage Switching Networks

By T. FELDMAN and J. W. RIEKE

Electronic switches have been applied in designs for multistage space-divided electronic switching networks. The electrical characteristics of the gas diode switch are discussed from the viewpoint of its utilization in switching networks, with emphasis upon the circuits employed in an experimental electronic switching system. A general method of calculating voltage margins for the reliable operation of multistage switching networks is illustrated by the description of two techniques for establishing a switching path, the end-marking and internal-marking methods. In conjunction with these operational methods, the uses of network controls and auxiliary network circuits, such as mark propagator and junctor circuits, are described. The transmission aspects of the electronic switching network determine to a large degree the circuit requirements and compromises. This is in contrast with the situation in electromechanical switching, where switching and transmission characteristics are largely independent.

I. INTRODUCTION

Switching networks, as discussed in this paper, comprise the interconnection facility for call traffic in a telephone switching system. This definition serves to exclude from consideration the numerous other components in the switching system, including the memory, logic, sensing and controlling elements that are required. This paper is concerned primarily with that large aggregation of switches which provides the desired voice frequency communication channels among telephone customers. Moreover, the discussion is limited to space-division electronic switching networks, excluding both electromechanical switching in its many forms and time-division electronic switching.

Since the advent of machine switching, a rapid evolution has taken place in the organization and control of multistage switching networks. This has led to the development of a succession of switching systems, each of which depended upon the nature of the chosen switching ele-

ment. In the Bell System these include step-by-step, panel and cross-bar systems; the system name in each case being derived from the type of switch employed. The electronic switching system currently under development is keyed by its title to that branch of technology which is so important a characteristic of the present era. Nevertheless, the properties of a new switch, the breakdown diode, again determine to a considerable degree the structure of the new system.[1] In this paper are described the breakdown characteristics of gas diode switches and the dc-coupled logic circuitry that is employed in the establishment of switching paths in a multistage breakdown electronic switching network. These techniques are general in their application. Also described are the control circuitry and transmission aspects of an electronic switching network.

II. GENERAL

The space-divided switching network is first of all a mapping problem. It takes the form of a network of nodes and branches, including nodes designated as terminals. In the branches of the network are switches which are closed in sequences to provide paths among designated sets of terminals. Many paths exist simultaneously, but at a given time no node (or branch) is associated with more than a single set of designated terminals. The last requirement is a statement that unrelated paths are not coupled with one another but that n-way connections among more than two terminals may be permitted.

The economics of switching network design have been studied in terms of minimization of the number of switches required.[2] The boundary conditions include the number of terminals and the amount and kinds of traffic offered. Resulting configurations assign switch sizes, number of switching stages, wiring patterns and methods of operation — everything intended to produce the best traffic performance for the least cost over a defined range of office sizes. Electronic switch elements permit the fullest exploitation of such studies since they may be organized into groups with a minimum regard for regularity or "natural size" of the group. Considerations which, in other switching systems, set the optimum size ranges for stepping switches, crossbar switches or panel assemblies have little relevance in electronic switching, where the switches are available as discrete units.

The structure of large networks has taken the form of multistage arrays of rectangular switch modules. In crossbar systems, for example, the switch module is a unit of 10 or 20 verticals (inputs) connecting

TABLE I — RELATIVE ECONOMY OF SEVERAL NETWORK
CONFIGURATIONS

Number of Stages	Switch Module Size	Terminals Each Side	Number of Diode Switches	Erlangs (Blocking = 1%)	Diode Switches per Erlang
3	75 x 75	5,625	1,265,625	4,280	296
4	22 x 22	10,648	937,024	4,550	206
5	20 x 20	8,000	800,000	4,930	162
6	10 x 10	10,000	600,000	4,250	143

with 10 horizontals (outputs). These 10-by-10 or 20-by-10 grids of crosspoints are organized into primary and secondary groups in an orderly wiring pattern, regularized for ease of control and administration. The electronic switching network may be organized in a similar structure. The basic switch module is a 10-by-10 grid with 100 crosspoints, each of which is supplied with a gas breakdown diode. The 10-by-10 switches are organized into primary, secondary and tertiary "frames" with a wiring pattern designed to maximize the fanout. The resulting "superframe" provides network building units with 1000 inputs, each of which reaches 1000 junctor terminals. Junctor links are employed to interconnect as many as ten originating superframes with an equal number of terminating superframes. The over-all switching network at maximum capacity has 10,000-by-10,000 terminations in six stages of 10-by-10 switch units, and contains 600,000 switch elements.

Further study to optimize network structures of this scale may indicate different choices for switch unit size, different number of stages and nonuniformity in structure. The study of structure is not the topic of this report, however, and it is mentioned only to establish the multi-stage fanout framework which constitutes the electrical circuit problem. Table I tabulates the calculated cost and performance for several large uniform grid networks in the region of 5000 erlang capacity. Lower cost in switches is achieved as the number of stages is increased.

III. THE BREAKDOWN DIODE SWITCH

Fig. 1 shows the significant voltage-current characteristics of the gas diode breakdown switch. It has been assumed that the device is to be operated in one dc polarity only, biased when conducting at a suitable sustaining value of dc, and that it will transmit speech signals which produce variations in the sustain current. It has been assumed further that a path may be extinguished from a single connected termi-
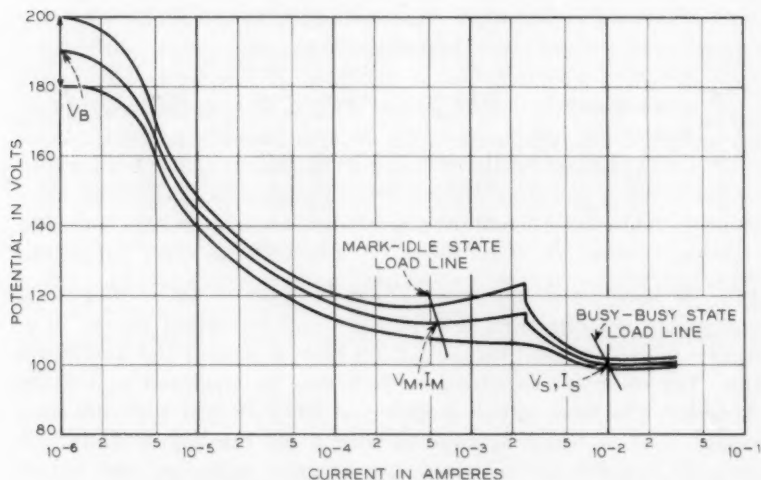
Fig. 1 — Breakdown characteristic for the A1624 gas diode switch in a multistage switching network application showing the range of product.

nal without a prior identification of the nodes which are associated in the particular path. This last assumption provides that the path will be extinguished if the normal sustain current is interrupted at any place in the path. Residual currents which might continue to flow through biasing impedances at the nodes must not sustain the switches.

The switch is made conducting by the application of a voltage at its terminals which exceeds the breakdown voltage, V_B with its range of values, $\pm \Delta V_B$. The circuit impedances establish the mark-idle load line shown in Fig. 1. The intersection of the load line with the switch characteristic established the value of current, I_M, which flows when the switch is marked. The voltage at this point, V_M, is lower than the marking voltage required to break down the switch. The voltage difference $V_B - V_M$ is the useful switch output signal which provides the mechanism whereby a marking potential, applied at the end of a switching network, propagates inward from stage to stage. The marking current, I_M, is held to a minimum practicable value since, in large switching networks, as breakdown propagates from the terminal to the interior nodes, this value is multiplied by fanout. Fanout ratios in large networks may exceed 1000-to-1, so that the total marking current requirement becomes impracticably large unless the current required to sustain the individual switch is maintained at a low value.

Those switches which are marked but not used in establishing a new connection must extinguish when the marking potential is again reduced to the idle value V_I, which is below the breakdown value. This also imposes restrictions upon the value of the marking current. The load line, when marking, must lie wholly above the switch characteristic; when idle, it must lie entirely beneath. Fig. 1 illustrates both situations for a typical switch. A third reason for minimizing the marking current is to maintain the load impedance as high as possible, since this is a shunt placed across busy paths in the talking condition and introduces audio losses. The final consideration bearing upon the value of the marking current is the negative impedance of the switch at this current: the lower the current, the higher the negative impedance. This impedance, together with the parasitic capacitance of the circuit, tends to produce sustained oscillations at marked nodes. Depending upon parasitic circuit capacitance to ground, marking currents in the order of 0.2 to 0.6 milliampere are typical. The V-I characteristic of the switch illustrates a typical discontinuity at $I = 2.5$ milliamperes. It is in this current region that the glow discharge of the gas tube moves from the outer surfaces of the cathode into its cavity.[3] Above this current, the tube continues to display a negative impedance—in this case, a useful transmission parameter. At the normal operating bias I_s, negative resistance of the order of 200 to 300 ohms is exhibited. The current in the chosen path is adjusted to the busy state (path-holding) value by changing simultaneously the load impedance and the source potential to path-holding values. The path-holding potential is below the voltage necessary to maintain conduction in the remaining marked switches, which are not required for the talking connection, and they extinguish. More will be said in a later part of the paper concerning the transmission properties of the gas diode switch and resulting network performance. At this point one other aspect is pertinent — the variabilities of all the switch parameters heretofore discussed. These ultimately determine the switching margins against false breakdown behavior. Successful design of a multistage network depends upon adequate limitation of the variabilities because, in ways which will become clear, they produce cumulative uncertainties in the process of propagating the marking potential into the fabric of a multistage switching network.

IV. BREAKDOWN SWITCHING NETWORKS

The crosspoint breakdown network exploits two useful characteristics of the switch element. One of these is that the path selection and con-

trol may be carried out over the path itself. The control voltages are applied to path nodes and directly operate the connecting switches. The second property of the switch is its utility as a memory cell. Its conducting or nonconducting state is read nondestructively in the path-finding process, thereby providing the mechanism for discerning idle nodes. No auxiliary path memory is required for these switching networks. Busy paths are avoided in the connection process without prior knowledge about the existing nodal busy or idle states. The nodal voltages shift potential when the node becomes busy, biasing the unconnected adjoining switches so that they cannot be broken down.

A node, q, in a multistage network may be in one of several states: busy, idle, marked or primed. The idle state potential is derived through shunt-bias impedances connected at the nodes. The prime-state potential is similarly derived but is applied in the marking sequence to designated nodes when directing the new paths to specified switch units, frames or switch levels is desired. The busy and marked potentials are derived from the various terminal circuits and transmitted through the switches to all the nodes in the respective marked or busy paths.

A marked node is expected to break down switches which connect with primed nodes in the succeeding stage. However, busy nodes are expected to be inaccessible from the adjacent nodes, excepting, of course, those which are associated with it in a busy path. These expectancies are met by a combination of properly chosen state potentials operating successfully with the voltage-current characteristics of the switches.

V. SWITCHING NETWORK BREAKDOWN MARGINS

One of the major design considerations in the construction of multistage switching networks is the provision of adequate voltage margins against false connections. A voltage margin may be defined as the minimum breakdown voltage of a switch less the maximum voltage that appears across it when it is in proper operation and should not break down. The following sections will illustrate a general method of calculating such margins in large networks and give some specific applications to several marking techniques. To facilitate the description of the network operation and margin conditions, the following list of terms, abbreviations and symbols is used:

n = number of network stages,

q = a general node in the network,

M_q^x = an x-type voltage margin between node q and a succeeding connecting node $(q + 1)$,

RM_q^x = an x-type voltage margin between node q and a preceding connecting node $(q - 1)$,

VI_q = voltage at node q during idle condition,

VK_q = voltage at node q during busy condition,

VM_q = voltage at node q during marking condition.

Fig. 2 shows two paths through an idealized n-stage network where there is no associated link circuitry. Only a general node q will be discussed. To provide adequate margins for the reliable operation of the network, it is necessary to select the worst possible condition contributing to false breakdown which exists between adjacent nodes, and then to optimize that condition. At a given time, a general node q may be in idle, busy or in some form of marking or priming condition. These potentials are indicated by the symbols VI_q, VK_q and VM_q, respectively. In addition, the busy node potential includes an ac component V_a, which represents the peak speech voltage transmitted through network paths. It is then possible to define the following set of margins to prevent breakdown of the switch in the forward direction.

(a) Margin M_q^a, which will prevent breakdown of a series switch when node q is busy, and node $(q + 1)$ is idle:

$$M_q^a = (V_B - \Delta V_B) - (VK_q - VI_{q+1}) - V_a. \qquad (1)$$

(b) Margin M_q^b, which prevents a series switch from breakdown when node q is busy and $(q + 1)$ is busy:

$$M_q^b = (V_B - \Delta V_B) - (VK_q - VK_{q+1}) - V_a. \qquad (2)$$

(c) Margin M_q^c, which will prevent breakdown when node q is busy and node $(q + 1)$ is in the marking condition:

$$M_q^c = (V_B - \Delta V_B) - (VK_q - VM_{q+1}) - V_a. \qquad (3)$$

(d) Margin M_q^d, which prevents breakdown of a switch when node q is idle and node $(q + 1)$ is also idle:

$$M_q^d = (V_B - \Delta V_B) - (VI_q - VI_{q+1}). \qquad (4)$$

(e) Margin M_q^e, which prevents a switch breakdown when node q is idle and node $(q + 1)$ is busy:

$$M_q^e = (V_B - \Delta V_B) - (VI_q - VK_{q+1}) - V_a. \qquad (5)$$

(f) Margin M_q^f, which will prevent a switch breakdown when node q is idle and $(q + 1)$ is in a marking condition:

$$M_q^f = (V_B - \Delta V_B) - (VI_q - VM_{q+1}). \tag{6}$$

(g) Margin M_q^g, which will prevent breakdown when node q is marking and $(q + 1)$ is idle:

$$M_q^g = (V_B - \Delta V_B) - (VM_q - VI_{q+1}). \tag{7}$$

(h) Margin M_q^h, which prevents a series switch breakdown when node q is in the marking condition and node $(q + 1)$ is busy:

$$M_q^h = (V_B - \Delta V_B) - (VM_q - VK_{q+1}) - V_a. \tag{8}$$

(i) Margin M_q^i, which prevents a switch breakdown when nodes q and $(q + 1)$ are in the marking condition:

$$M_q^i = (V_B - \Delta V_B) - (VM_q - VM_{q+1}). \tag{9}$$

This list of margins may be reduced if it be assumed that two adjacent nodes in the same condition do not provide sufficient potential across the intervening switch to cause a breakdown. This eliminates margins M_q^b and M_q^d as trivial. It should be noted that, when there is more than one form of marking involved in the operational technique, the number of margins will increase. Under conditions which depend upon the particular network configuration, certain members of the preceding list may become the necessary conditions for the desired breakdown of the switch. In these cases, the voltage differences become negative, indicating the assured breakdown of the switch. An example of this would be condition M_q^i, which assures forward breakdown along the path of the desired connection.

A similar number of margins may also be defined in the reverse direction, i.e., voltage conditions which operate the switch in their reverse breakdown condition. As an example, RM_q^a may be defined as the margin preventing a reverse breakdown of a device when node q is in the busy condition, and node $(q - 1)$ is idle, i.e.:

$$RM_q^a = (V_{RB} - \Delta V_{RB}) - (VK_q - VI_{q-1}) - V_a. \tag{10}$$

Combining (1) and (10) gives a comparison of forward and reverse margins:

$$RM_q^a - M_q^a = (V_{RB} - V_B) - (\Delta V_{RB} - \Delta V_B) + (VI_{q-1} - VI_{q+1}). \tag{11}$$

It is characteristic of the gas diode network that the margins in the reverse direction exceed the forward margins. In gas diode networks the

sustaining voltage drop of the device is substantial (approximately 100 volts). To permit the breakdown of successive switches in establishing a connection, it is necessary to have a stage-to-stage increment in the idle node voltage bias about equal to the sustaining voltage drop. In (11) the term $(VI_{q-1} - VI_{q+1})$ is then approximately 200 volts, assuring a large reverse breakdown margin.

For devices having a very low sustaining voltage drop in the forward direction, the term to be considered in (11) is $(V_{RB} - V_B)$. To achieve a positive excess reverse breakdown margin, it is necessary that the switch reverse breakdown voltage, V_{RB}, safely exceed the forward breakdown, V_B.

Comparisons may be made among the other pairs of forward and reverse margins as was done in (11). Such comparisons lead to the same result, namely, the reverse margins to prevent breakdown exceed corresponding forward breakdown margins. The poorest of the reverse margins sets the requirement for the switch reverse breakdown voltage. Except for this calculation of reverse breakdown requirement for the switch, consideration now need be given only to the margins in the forward direction.

The list of margins may now be reduced by inspection. Assume, as a reference voltage, the busy condition at q to be VK_q. Then the idle and marking conditions may be either positive or negative with respect to this reference potential. Combining equations (1) and (5) gives, for example,

$$M_q^a - M_q^e = (VI_q + VI_{q+1}) - (VK_q + VK_{q+1}).$$

Then, depending on the relative polarity of the idle condition, M_q^e is greater than or less than M_q^a. In a particular network only one condition will exist, and consequently only one of M_q^a or M_q^e need be considered. Similarily, it can be shown that only one of M_q^c or M_q^h and only one of M_q^f or M_q^g need be considered. If the list of possible potentials for the idle and marking conditions is then considered with respect to the busy voltage as a reference, Table II may be drawn up (for convenience, the subscript q is dropped).

The worst marginal set now being known, it is necessary to optimize within the set in order to determine the best values of the circuit parameters. As previously indicated, the negative margin M_q^i, which assures forward breakdown along the desired path, is combined with the set of worst margins and provides a system of simultaneous equations in the bias voltages which may be solved for the particular case.

This will be more clearly understood if an example is considered. As

TABLE II — NETWORK MARGINAL SETS

VK	VI	VM	$\mid VM \mid >$ $\mid VI \mid$	$\mid VI \mid >$ $\mid VM \mid$	To Be Considered	To Be Eliminated
Reference	+	−			$M^f\ M^e\ M^c$	$M^a\ M^g\ M^h$
Reference	−	+			$M^g\ M^a\ M^h$	$M^e\ M^f\ M^c$
Reference	+	+	Yes	No	$M^h\ M^e\ M^g$	$M^a\ M^c\ M^f$
Reference	+	+	No	Yes	$M^e\ M^h\ M^f$	$M^a\ M^c\ M^g$
Reference	−	−	Yes	No	$M^c\ M^a\ M^f$	$M^e\ M^h\ M^g$
Reference	—	—	No	Yes	$M^a\ M^c\ M^g$	$M^e\ M^h\ M^f$

an application of margin calculations, consider a network employing a technique of end, or terminal, marking[4] to establish a connection.

VI. END-MARKING TECHNIQUE

This technique is a method of controlling the establishment of connections in a switching network where voltage (or impedance) marks are derived by translation of address information and applied to network terminals only.

The propagation of the marks into the fabric of the network can be described briefly as follows. In the idle condition, all nodes are biased negatively. To establish a connection, a terminal is selected and a marking potential, VM_0, is applied to the zero, or terminal, node. This is a positive potential and, combined with the negative idle voltages appearing at the connecting first nodes, it is sufficient to switch the first-stage devices to their low-impedance state. Associated with a switch in its busy state is a potential drop, V_s, and its variation, ΔV_s. The potential drop, $V_s \pm \Delta V_s$, now occurs across one of the first-stage crosspoints which switches to the low-impedance state, and a voltage $VM_0 - (V_s \pm \Delta V_s)$ appears at the first nodal position. This, combined with the negative idle potential at the second nodes, is sufficient to break down
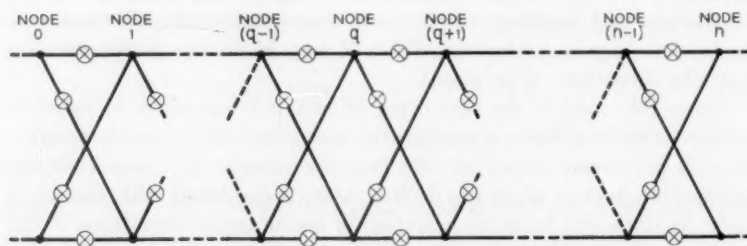

Fig. 2 — Two paths through an n-stage switching network.

the connecting second-stage switches. Multipled at each node are several paths leading through second-stage switches to secondary nodes; two such paths are shown in Fig. 2. The idle potential at any such multipled second nodes will cause breakdown of additional second-stage switches, permitting the marking potential to fanout to several switches in the second stage. It is this fanout of the marking potential which provides the ability of the network to seek out idle paths. The marking potential at one node tries to break down all multipled switches; those which do break down form elements of possible paths. By this method, the terminal mark produces progressive breakdown towards the outgoing terminal. Since sufficient current must be provided to maintain the multipled switches in the second (and ensuing) stages in their low-impedance states, it is readily seen that an important inherent drawback of the end-marking technique is the very large current drain. (In three stages of 10 x 10 switches, this fanout ratio is 1,000-to-1.)

As a result of the voltage drops across switches in the busy condition, the busy voltage, VK_q, at a node q depends upon the position in the network at which the reference voltage in the busy path, V_R, is introduced. For simplicity in discussion, assume that the dc reference voltage, $-V_R$, is applied at the outgoing terminal, and has associated with it a small variation, ΔV_R. The busy voltage at node q is then given by the relation

$$VK_q = -(V_R \pm \Delta V_R) + (n - q)(V_S \pm \Delta V_S). \tag{12}$$

Experimentally, it has been found that, to switch a breakdown diode to its low-impedance "on" state, it is necessary to provide a voltage greater than the nominal breakdown voltage and its variation, $(V_B + \Delta V_B)$. This overvoltage is given by the symbol Δ. Then, for the end-marked network, the marking potential at a node with idle potential at the next adjacent node must turn the switch "on," i.e.

$$VM_q - VI_{q+1} = (V_B + \Delta V_B + \Delta). \tag{13}$$

Equation (13) represents condition (g) of the preceding margin list, where the voltage difference is negative, assuring the breakdown of the switch. Moreover, the difference in the marking potentials at adjacent nodes is merely the sustaining voltage (and variation) of the intervening switch, i.e.,

$$VM_q = VM_{q+1} + (V_S \pm \Delta V_S), \tag{14}$$

or, in general,

$$VM_q = VM_{q+x} + x(V_S \pm \Delta V_S). \tag{15}$$

A tentative assignment of potentials can now be made for VM_q and VI_q. As indicated in (15), the marking potential changes from node to node decreasing as q increases, and cannot be set to a universal value. The idle potential at all nodes may, as a practical matter, be set to a single value. The marking potential has its least value at node $(n - 1)$. Then, to satisfy (13),

$$VM_{n-1} + VI_n = (V_B + \Delta V_B + \Delta), \qquad (16)$$

and VI_n will represent the largest idle potential in the network. All other idle potentials will be less than this, and any universal idle potential must at least equal this value of VI_n. Choose, then, all idle potentials to be the same as VI_n, and set

$$VI_n = VI_q = -(1 - \alpha)(V_B + \Delta V_B + \Delta), \qquad (17)$$

where α is a fraction less than, or equal to, unity. From (16) and (17),

$$VM_{n-1} = \alpha(V_B + \Delta V_B + \Delta).$$

From (15),

$$VM_q = VM_{n-1} + (n - q - 1)(V_S \pm \Delta V_S).$$

Combining these two yields

$$VM_q = \alpha(V_B + \Delta V_B + \Delta) + (n - q - 1)(V_S \pm \Delta V_S). \quad (18)$$

Then, for the end-marked network, (15), (17) and (18) give the values of the busy, idle and marking voltages. To determine the worst marginal conditions, the busy voltage is taken as a reference voltage, and these equations then represent the conditions in the second row of Table II. The margins which must be considered are M_q^a, M_q^g, and M_q^h. The conditions given by M_q^g have already been considered in obtaining (13). The two remaining conditions which must be studied are

$$M_q^a = (V_B - \Delta V_B) - (VK_q - VI_{q+1}) - V_a$$

$$= \alpha V_B - (2 - \alpha)\Delta V_B + V_R - \Delta V_R$$

$$- (1 - \alpha)\Delta - (n - q)(V_S + \Delta V_S) - V_a,$$

$$M_q^h = (V_B - \Delta V_B) - (VM_q - VK_{q+1}) - V_a$$

$$= (1 - \alpha)V_B - (1 + \alpha)\Delta V_B - V_R$$

$$- \Delta V_R - \alpha \Delta - 2(n - q - 1)\Delta V_S - V_a.$$

Both these margins exhibit the worst conditions at the same node,

$q = 0$. To optimize conditions, these two margins may be equated, yielding, at $q = 0$,

$$2V_R = (1 - 2\alpha)(V_B + \Delta V_B + \Delta) + nV_S - (n - 2)\Delta V_S. \quad (19)$$

From this equation it may be seen that the reference voltage, V_R, is most independent of device parameters if $2\alpha = 1$. Under these conditions,

$$2V_R = nV_S - (n - 2)\Delta V_S. \quad (20)$$

The worst margin may then be expressed as

$$M_0{}^a = M_0{}^h = \frac{1}{2} V_B - \frac{3}{2} \Delta V_B - \frac{1}{2} \Delta - \Delta V_R - \frac{n}{2} V_S$$
$$- \left(\frac{3n - 2}{2}\right) \Delta V_S - V_a. \quad (21)$$

The overvoltage, Δ, may also be considered as one of the necessary margins in the network operation. Then, setting Δ equal to the other margin, the worst margin becomes

$$M = \frac{1}{3} V_B - \Delta V_B - \frac{2}{3} \Delta V_R - \frac{n}{3} V_S - \left(\frac{3n - 2}{3}\right) \Delta V_S - \frac{2}{3} V_a. \quad (22)$$

To determine the number of stages of switching through which the marking potential may be propagated, M is equated to zero, and (22) is solved for n. This gives

$$n = (V_B - 3\Delta V_B - 2V_R + 2\Delta V_S - 2V_a)/(V_S + 3\Delta V_S). \quad (23)$$

As a numerical example, suppose that the switch used is a neon gas diode whose characteristics are:

$V_B = 190 \pm 10$ volts,

$V_S = 101 \pm 1$ volt at 10 milliamperes sustain current,

$\Delta V_R = 1.0$ volt.

This network would operate in a 2000-ohm impedance circuit in order to transmit 10 milliwatts of audio power, and $V_a = 6.3$ volts. Substituting these values in equation (23) yields

$$n_{\text{neon diode}} = (190.0 - 30.0 - 2.0 + 2.0 - 12.6)/(101.0 + 3.0)$$
$$= 1.3.$$

That is, the marking potential may be transmitted reliably through just

one stage of switching; with two stages of switching, the margins become negative and false breakdowns will occur.

Various means for improving the margins for a gas diode network have been proposed. An obvious improvement is to bisect the network, so that it is marked at both terminals and these marks produce progressive breakdown toward the central nodes. At the center, a matching facility is necessary which will recognize the arrival of a clear path to both terminals and connect these paths. This matching device also permits the establishment of one connection only, which provides a lock-out function against redundant parallel paths. Such a bisection should approximately double the number of stages for which the use of a given switch is applicable. The network is symmetrical about the central node (junctor) and, consequently, the operation of only one-half the network need be studied. It is also possible to regulate the holding voltage, V_R, at the central position in the network, thereby minimizing the variation of the worst busy link voltage.

The busy voltage at a general node q is given by the relation

$$VK_q = -V_R + \Delta V_R + \left(\frac{n}{2} - q\right)(V_S \pm \Delta V_S). \qquad (24)$$

The propagation of the marking potential through the network is still represented by (13) to (15) and the marking potential has its least value at node $(n/2) - 1$. Then, to satisfy (13),

$$VM_{n/2-1} + VI_{n/2} = (V_B + \Delta V_B + \Delta).$$

The potential $VI_{n/2}$ represents the largest idle voltage, and all others may be chosen equal to it, i.e.,

$$VI_{n/2} = VI_q = -(1 - \alpha)(V_B + \Delta V_B + \Delta). \qquad (25)$$

Thence,

$$VM_{n/2-1} = \alpha(V_B + \Delta V_B + \Delta),$$

and, from (15),

$$VM_q = VM_{n/2-1} + \left(\frac{n}{2} - q - 1\right)(V_S \pm \Delta V_S).$$

Combining these two yields

$$VM_q = \alpha(V_B + \Delta V_B + \Delta) + \left(\frac{n}{2} - q - 1\right)(V_S \pm \Delta V_S). \qquad (26)$$

Then, for the bisected, end-marked network in which the reference voltage supply is placed at the junctor, the busy, idle and marking potentials are given by (24) to (26). These represent the conditions in the second row of Table II and, as in the previous example, the margins M_q^a, M_q^g and M_q^h must be considered. Once again, M_q^g represents the condition for the desired forward fanout of the marking voltage. The other margins may be written as:

$$M_q^a = (V_B - \Delta V_B) - (VK_q - VI_{q+1}) - V_a = \alpha V_B - (2 - \alpha)\Delta V_B$$

$$+ V_R - \Delta V_R - (1 - \alpha)\Delta - \left(\frac{n}{2} - q\right)(V_S + \Delta V_S) - V_a,$$

$$M_q^h = (V_B - \Delta V_B) - (VM_q - VK_q + 1) - V_a = (1 - \alpha)V_B$$

$$- (1 + \alpha)\Delta V_B - V_R - \Delta V_R - \alpha\Delta - 2\left(\frac{n}{2} - q - 1\right)\Delta V_S - V_a.$$

Both these margins exhibit their worst condition at the node $q = 0$. To optimize conditions, these margins are equated at $q = 0$, yielding

$$2V_R = (1 - 2\alpha)(V_B + \Delta V_B + \Delta) + \left(\frac{n}{2}\right)V_S - \left(\frac{n}{2} - 2\right)\Delta V_S. \quad (27)$$

As before, V_R is most independent of device parameters if $2\alpha = 1$ and, under these conditions,

$$2V_R = \left(\frac{n}{2}\right)V_S - \left(\frac{n}{2} - 2\right)\Delta V_S, \quad (28)$$

and the margins become

$$M_0^a = M_0^h = \tfrac{1}{2}V_B - \tfrac{3}{2}\Delta V_B - \tfrac{1}{2}\Delta - \Delta V_R - \frac{n}{4}V_S$$

$$- \left(\frac{3n + 4}{4}\right)\Delta V_S - V_a. \quad (29)$$

If the overvoltage is considered a necessary margin and is equated to the margin in (29), the worst margin is

$$M = \tfrac{1}{3}V_B - \Delta V_B - \tfrac{2}{3}\Delta V_R - \frac{n}{6}V_S - \left(\frac{3n + 4}{6}\right)\Delta V_S - \tfrac{2}{3}V_a. \quad (30)$$

Then, to determine the number of stages of switching through which the marking potential may fan out without a negative margin, (30) is equated to zero, and

$$n = 2(V_B - 3\Delta V_B - 2\Delta V_R + 2\Delta V_S - 2V_a)/(V_S + 3\Delta V_S). \quad (31)$$

This equation gives n as exactly twice the value for a nonbisected network in which V_R is supplied at the outgoing terminal, as indicated in (23).

Then, for a gas diode switch, the bisected network can permit the propagation of the marking potential through two stages of switching before false breakdowns will occur.

Although there will be no change in the margin equations just derived, it is possible to obtain a more symmetrical arrangement of the network. The switches in each half of the bisected network may be polarity-oriented towards the junctor. This symmetrical arrangement exhibits better transmission properties for the network than did the preceding "straight-through" case. Gas diodes of the type presently available are used in a region of nonlinear negative resistance. In the straight-through network, instability and distortion arise from this nonlinearity.

For a bisected network, then, it would appear that a maximum number of stages would be two, unless additional control of the circuitry were added.

VII. PROPAGATORS

Such an additional circuit is the propagator. This is an auxiliary link circuit which, in a multistage network, would appear at intermediate nodes. The propagators correct for the variations in switch characteristics by requantizing the voltage mark as it is transmitted towards the junctor. This regenerated potential can then fan out through two additional stages and still maintain positive margins.

A propagator circuit, for use with a gas diode switch, is shown in Fig. 3. A marking signal, in the form of a positive pulse, ionizes the gas diode, X_1, in the stage of switching preceding the propagator. An idle input bias and link resistor provide that, when X_1 ionizes, current flows from X_1 to the idle input bias terminal. When X_1 breaks down to its low-impedance state, the voltage at the cathode of X_1 will increase by the difference in the sustaining and idle voltages of the diode switch. The increase is ac-coupled through a resistor-capacitor combination, R_1-C_1, to the anode of the propagator gas diode. The diode D_2 is reverse-biased, and the voltage coupled through R_1-C_1 will add to the idle voltage already appearing at the propagator anode. When the positive pulse at the cathode of X_1 rises above the potential of the idle output bias, the diode D_1 becomes forward-biased and a small part of this pulse is reflected through D_1 to the cathode of the propagator tube. Despite the appearance of part of the pulse at the propagator cathode, the main

part of the pulse arrives at the propagator anode. The pulse reaching the anode of the propagator has an exponential decay with a time constant $C_1(R_1 + R_2)$.

To assure breakdown of the propagator tube, this pulse must remain above a given breakdown voltage for a time greater than the longest ionization time of the tube. Simultaneously with the application of marking voltages, a propagator mark voltage pulse appears at all propagators. This rises to the same potential as the idle propagator bias, thereby giving this supply a low-impedance path on the anode side of the propagator tube. In the busy paths the propagator gas diodes are not ionized by the propagator mark voltage. This voltage is coupled into the circuit through diode D_2 at a magnitude which does not exceed the propagator idle bias, so that the marking pulse will not be coupled into busy nodes through the R_1C_1 network. The source impedance of the propagator mark voltage is low so that, upon ionization of the propagator gas diode, a new source of marking potential is effectively coupled to the output node. The cathode voltage rises to the clamping potential, the mark limiting voltage. This is the newly regenerated marking potential which propagates forward into the succeeding switching stage X_2.

The propagator also reduces the fanout current for the X_1 stage of the distribution network during the marking cycle. This, of course, reduces the current requirement of the voltage-marking source at the network terminal.

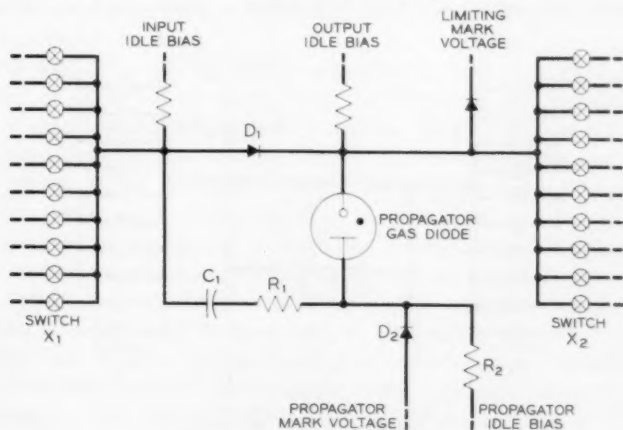With propagators, a six-stage network may be constructed to operate

Fig. 3 — Active mark propagator circuit.

with positive voltage margins. The four inner stages, bisected as already described, now operate as two two-stage networks. Between these and the two outer stages are inserted propagator circuits. In this arrangement each half of the network is marked as a one-stage network and a two-stage network and is held, in the busy condition, as a three-stage network. The margins lie between those of a two- and a three-stage network; examination shows them to be positive and close to those of the two-stage network alone.

The propagator just described operates to connect additional marking energy into a switching node and provides the dual results of requantizing the mark potential and supplying the fanout current. When current fanout is not a consideration, the following passive circuit (Fig. 4) may be employed at less cost. This circuit is useful, for example, between stages two and three and between stages four and five of the six-stage network as an additional margin improvement circuit. No additional mark voltage source or active device is required.

In the idle condition, diode D_3 is reverse-biased, and a charge appears across capacitor C_2. When a connection is being established, a marking signal appears at the anode of the second-stage switch X_2. This, combined with the input idle bias, is sufficient to ionize X_2, and the voltage at the cathode of X_2 rises by the difference in the sustaining and idle voltages of the gas diode. This voltage pulse is not transmitted through D_3 to the anode of the third-stage switch X_3. The pulse is, however, ac-coupled through the R_3-C_2 resistor-capacitor network. The pulse combines with the output idle bias to produce a pulse which is limited by
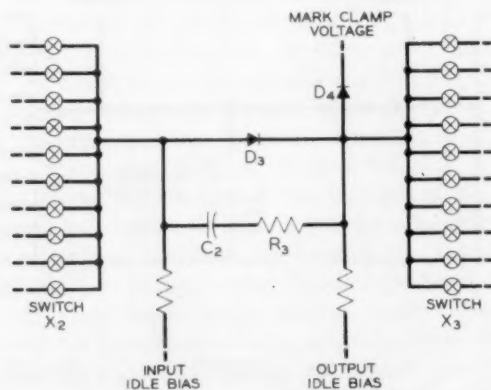

Fig. 4 — Passive mark propagator circuit.

conduction of diode D_4 to the clamp mark voltage. This voltage is subject to less variation than the original marking signal appearing at the anode of the second-stage switch.

The requantized marking voltage is supplied to the anode of the third-stage switches X_3 and, combined with the third-stage output idle bias, it is sufficient to ionize the X_3 switches. The discharge time of the R_3-C_2 network must be less than the time required to ionize the X_3 switches, so that the potential appearing at the input node to stage three does not come out of the clamp before the X_3 switches ionize. When the X_3 switches conduct, the potential at the X_3 node falls, and diode D_3 then conducts in the forward direction, completing a dc path from the second to the third stage of the network.

VIII. NETWORK CONTROLS

To this point the discussion primarily has been concerned with the distribution switching network; there has been a description of the end-marking process, the switch characterization and circuits such as the propagator which implement the marking process. There are, however, several types of orders which must be carried out by the network.[1] Information must flow between the memory elements of the system and various network controls which carry out these orders. There must also be a sequence circuitry for programming these orders through the network. The following sections will briefly discuss network controls and their operation such as those which will be used in the experimental electronic switching system.

Connections are set up in the distribution switching network by marking, or changing the voltage, at the particular end terminals between which a transmission path is desired. This change of voltage is propagated through the network from each side toward the center on all idle paths. At the junctor only one of the many paths that can handle the call is permitted to connect itself to the hold supply.

To establish a connection through the network, the binary address of the desired terminal is sent from the central control of the system to the input of a selector circuit. The selector contains two stages of magnetic core circuitry. It translates in two stages the given address to a two-digit address, one 20-valued and one 50-valued. Seventy leads then are wired through a maximum of 1000 terminals. Each terminal has a resistor and a selector gas diode, which act as the last stage of selection. These elements are wired to the 20-by-50 crosspoint matrix. A negative pulse on one of the 50 X-coordinate leads and a positive

pulse on one of the 20 Y-coordinate leads will cause one, and only one, of the network terminal selector gas diodes to ionize. This transmits the marking potential to the anodes of the first-stage switches.

Associated with the selector is a sequential pulsing circuit. A pulse into this circuit clears both stages of the selector core circuitry. The input cores then are set according to the voltage pattern on the address leads. Finally, the information is read through the two stages of cores and fires one selector diode in the output stage.

At the end of this operation, three pieces of information are returned to the central control of the system: (a) the completion of the operation, (b) its success or failure, and (c) for an unsuccessful connection, whether a failure was caused by encountering a busy terminal or by failing to make a connection. This latter case may be caused by a fault in the system or by inability to find an idle path from an available junctor to the desired terminal, i.e., blocking.

When the desired terminal address is sent to the selectors, other operations are also begun. The propagator pulsers are turned on, and the "time-out" pulser is set. This pulser is a clock which measures the interval permitted in which to establish a connection.

The busy-check delay circuit is also set in operation. This is a mono-stable flip-flop which times the check of end terminals for busy connections. If either of the end terminals is busy within the check period, the junctor enabler is not started and the operation is concluded.

If both terminals are idle, the junctor enabler circuit, which sequentially tests junctors for the presence of mark signals, is set in operation. A junctor circuit and control are shown in Fig. 5. A junctor latch tube receives marks from both sides of the network and, with the enabling signal, it ionizes and completes the paths from the marked terminals through the two halves of the network to the holding potential. A match pulse is generated by the resulting holding current which indicates the successful completion of the connection.

If, at the end of the permitted "time-out", neither a "successful match" nor "busy" signal is received, the time-out pulser reports this to the central control and the network controls are cleared.

The controls are cleared after each operation of a sequential circuit. A pulse from the match (or release) detector, from the connect (or release) time-out pulser, or from the central control system can initiate such a clearing sequence. In the connect sequence, a pulse from any of these sources stops the junctor enabler and resets the time-out pulsers. Deionization of the gas tubes in the output stages of the propagator

pulsers and selectors begins, and the control voltages are removed from the network.

After sufficient time has elapsed to clear the network and its controls, a pulse is transmitted to the central control which signals the readiness of the network to accept the next command.

To disconnect a path, the address of either terminal of the connection is fed from the central control to the appropriate terminal selector. This translates the address in the same way as for the connect order and reapplies the mark voltage at the specified terminal. Associated with the release order is a delay circuit which, upon operation, initiates the "release time-out" clock. The mark voltage is transmitted via the connection to the junctor. This, in conjunction with a release pulse applied in common to all junctors, fires a release tube in the junctor which removes the hold current from the latch tube. When the current in the latch tube has decreased below its minimum sustaining current, deionization takes place. The release tube then is extinguished by removal of the common release pulse, and the network release is complete.
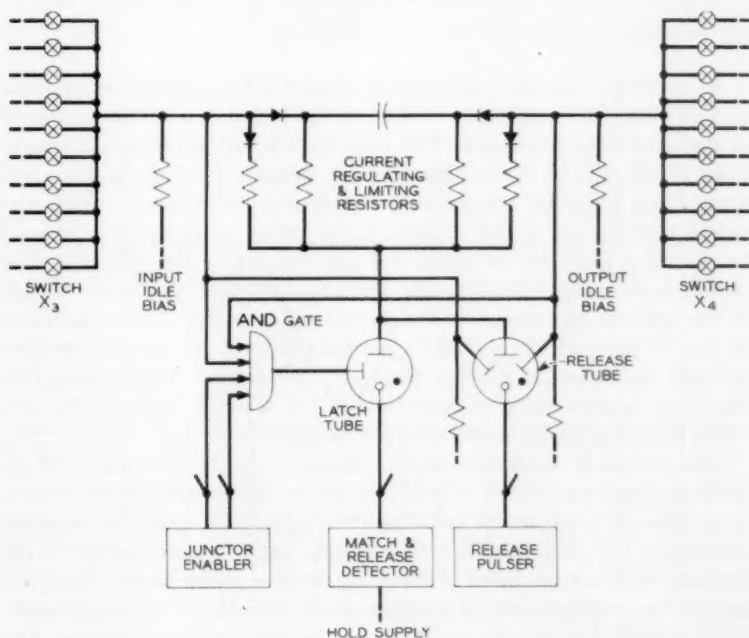

Fig. 5 — Junctor circuit and contro .

Network control circuits include one more group, called "identifiers." These are the inverse of selectors; they detect at network terminals changes in current which are produced by marking or release actions and translate the specific action location into its binary address. This function serves two important purposes in network operation. First, the operation of the selectors which translate the input binary address into action at a specific terminal is checked by the simultaneous inverse translation of the action back to binary address form. The selector address is compared with the identifier address. If the two binary numbers match, it is certain that marking potential was connected to the specified terminal. Mismatch is a valuable trouble indication and a signal to operate with the duplicate set of network control equipment.

The second use of the identifier is to permit a trace through the network. When a disconnect is initiated at one terminal in a connection, the equipment number of the opposite terminal is recovered in its corresponding identifier. Progressive disconnect is thereby facilitated without the need for large amounts of call memory.

IX. INTERNAL MARKING

An alternative to the end-marking scheme is to program the marking of each stage of switching according to the terminal addresses involved in the particular connection. The network terminal address is a binary word which may be divided into several sections. For the six-stage network, these sections specify the appropriate superframe, frame and switch components of the address. Each section of the address is used to control a sub-selector and a full selector marks the terminal. The switch selector has access to the first nodal circuits; the frame section of the address has access to the second set of nodes. The superframe section of the address is used to partially control the junctor enabler, limiting its access to junctor nodes in the selected superframe. This technique applies marking potentials to the selected nodes in the network, directing the establishment of connections.

The system is characterized by the addition, at each node, of an auxiliary marking circuit which includes a shunt breakdown switch. It is through these shunt switches that individual nodes are marked, permitting the establishment of a unique path in both halves of the network. A selective choice of matching devices at the junctor then will permit the establishment of a unique path through the entire network. This procedure is similar to that used in the crossbar systems. The ability to determine a single switching path eliminates the problem of

fanout that occurs in the end-marked network, giving a considerable saving in marking-current drain.

Two paths through one half of such an internally-marked n-stage network are shown in Fig. 6. The network is bisected and the reference voltage in a busy path, $-V_R$, is introduced at the central node. The operation of the system is symmetrical about the junctor, and only that of one half need be considered.

As for the bisected end-marked network, the busy voltage at a general node, q, is given by the relation

$$VK_q = -(V_R \pm \Delta V_R) + \left(\frac{n}{2} - q\right)(V_s \pm \Delta V_s). \tag{32}$$

The sequence employed for establishing a connection is now detailed. In the idle condition the nodes are permanently biased through a resistor, R_q, at a negative potential, E_q. When a node has been selected to participate in the establishment of a connection, a positive potential, V_q, is applied through the shunt switch in series with another resistor, r_q. The voltage, $V_q - E_q$, is sufficient to operate the shunt switch at a sustaining current, $I_p \pm \Delta I_p$, somewhat in excess of its minimum sus-
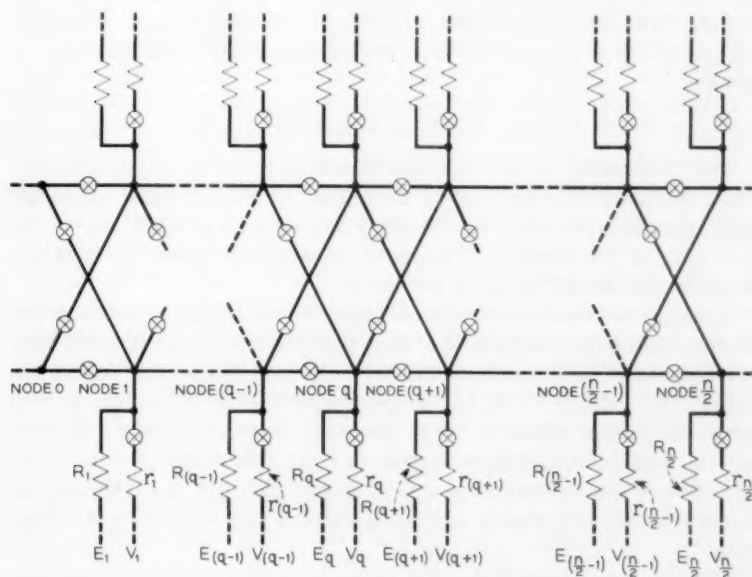

Fig. 6 — Two paths through $n/2$ stages of internally marked switching network.

taining current, i.e.,

$$V_q - E_q = (V_B + \Delta V_B + \Delta). \tag{33}$$

If the node is already busy, the application of V_q will not operate the shunt switch. Thus, busy paths are not disturbed by the marking of nodes in the process of setting up new paths. This change of state at the node will be called priming, and the potential at q during this condition is VP_q.

The junctor node, in the idle condition, is permanently primed at a value $VP_{n/2}$ through the resistor $R_{n/2}$. If an idle junctor node is selected, a negative potential $V_{n/2}$ is applied through $r_{n/2}$ to turn on the shunt switch at the node $n/2$. This places a negative potential $VMI_{n/2}$ at the node, where $VMI_{n/2}$ is called the first marking potential at node $n/2$. If now a connecting preceding node is in the primed condition, the series switch between these nodes will turn on, i.e.,

$$VP_{(n/2)-1} - VMI_{n/2} \geqq (V_B + \Delta V_B + \Delta). \tag{34}$$

The potential now appearing at node $n/2 - 1$ is $VMI_{n/2-1}$. This negative potential is such that, combined with the priming potential at the next preceding connecting stage, it will turn on the intervening switch. To assure this condition, $r_{n/2} \ll r_{n/2-1}$. In this manner, the marking potential is propagated outwards towards the terminals and, in general,

$$VP_{q-1} - VMI_q \geqq (V_B + \Delta V_B + \Delta). \tag{35}$$

The resistances, r_0 and $r_{n/2}$, are such that $r_0 \ll r_{n/2}$. Consequently, when the mark has reached the zero node (terminal), nodal potentials again reverse, and a positive potential now appears at each node in the first half of the network. This positive potential, given by $VMII_q$, is called the second marking potential.

A similar succession of potentials appears at the nodal positions in the second half of the network. If a clear path for the connection has been encountered in both halves of the network then, during the junctor enabling, the potential $VMII_{n/2}$ coupled with the enabling voltage will cause the junctor switch to break down and conduct in the forward direction. This permits a large current to flow in the series switches. The detection of this current surge sets in motion circuitry to extinguish the shunt switches and supply a holding current and voltage in the busy path.

The priming voltage, VP_q, at a node q is determined by the bias voltage applied to the node, V_q, i.e.,

$$VP_q = V_q - r_q(I_p + \Delta I_p) - V_s - \Delta V_s. \tag{36}$$

Combining (35) and (36) yields

$$-VMI_q = V_B + \Delta V_B + \Delta$$
$$+ r_{q-1}(I_p + \Delta I_p) + V_s + \Delta V_s - V_{q-1}. \tag{37}$$

A tentative assignment of potentials may be made for V_q. This may have the same value at all nodes, and so may be set to

$$V_q = \beta(V_B + \Delta V_B + \Delta), \tag{38}$$

where β is a fraction less than, or equal to, unity. Then, to satisfy (33),

$$E_q = -(1 - \beta)(V_B + \Delta V_B + \Delta). \tag{39}$$

The network operation involves three forms of marking and, as a consequence, the number of margins to be considered is increased three-fold over the number considered when one form of marking was employed, as in the end-marked case. That is, VM of Table II now has three possible values, VP, VMI and $VMII$. The potential at node q under differing conditions is given by equations (32) to (39). Using the busy potential at node q as a reference, the idle and priming voltages represent the conditions of line two of Table II. The idle, busy and first marking voltage give the conditions of line five of Table II. Since there is some freedom in the choice of the junctor enabling potential, $VMII_{n/2}$ may be chosen so as to give better-than-minimal margin conditions. For example, $VMII_{n/2}$ may be chosen to be $\frac{1}{2}(V_B + \Delta V_B + \Delta)$. Then the idle, busy and second marking voltage give the conditions on line two of Table II. These conditions, however, are less restrictive than those represented by the idle, busy and priming voltages. The marginal conditions to be considered, then, are (a), (c), (f), (g) and (h) of Section V.

Marginal conditions (f) and (g) are merely restatements of the conditions necessary for the proper network operation. An examination of the remaining margins shows that M_q^a may be eliminated as being less restrictive than M_q^c and M_q^h. Then the two worst conditions are given by:

(i) Margin M_q^c, which prevents the breakdown of a series switch when node q is busy and node $(q + 1)$ is in the first stage of marking, i.e.,

$$M_q^c = (V_B - \Delta V_B) - (VK_q - VMI_{(q+1)}) - V_a.$$

From (32), (37) and (38), this becomes

$$M_q^c = \beta V_B + (\beta - 2)\Delta V_B + (\beta - 1)\Delta + V_R - \Delta V_R$$
$$- r_{c-1}(I_p + \Delta I_p) - \left(\frac{n}{2} - q + 1\right)(V_s + \Delta V_s) - V_a. \tag{40}$$

(ii) Margin M_q^h, which prevents the breakdown of a shunt switch when node q is in the busy condition and is subsequently primed, i.e.,

$$M_q^h = (V_B - \Delta V_B) - (V_q - VK_q) - V_a.$$

From (32) and (38) this becomes

$$M_q^h = (1 - \beta)V_B - (1 + \beta)\Delta V_B - \beta\Delta - V_R - \Delta V_R \\ + \left(\frac{n}{2} - q\right)(V_S - \Delta V_S) - V_a. \tag{41}$$

Margin M_q^c exhibits its worst conditions at node $q = 1$; M_q^h has its worst conditions at $q = n/2$. To optimize conditions, M_1^c is equated to $M_{n/2}^h$, yielding,

$$2V_R = (1 - 2\beta)(V_B + \Delta V_B + \Delta) \\ + \left(\frac{n}{2}\right)(V_S + \Delta V_S) + r_0(I_p + \Delta I_p). \tag{42}$$

As before, V_R is most independent of device parameters if $2\beta = 1$ and, under these conditions,

$$2V_R = \left(\frac{n}{2}\right)(V_S + \Delta V_S) + r_0(I_p + \Delta I_p), \tag{43}$$

and the worst margin becomes

$$M_1^c = M_{(n/2)}^h = \tfrac{1}{2}V_B - \left[\tfrac{3}{2}\Delta V_B + \tfrac{1}{2}\Delta + \Delta V_R + \tfrac{1}{2}r_0(I_p + \Delta I_p) \\ + \frac{n}{4}(V_S + \Delta V_S)\right] - V_a. \tag{44}$$

If the overvoltage is considered a necessary margin and is equated to the margin in (44), the worst margin is

$$M = \tfrac{1}{3}V_B \\ - \left[\Delta V_B + \tfrac{2}{3}\Delta V_R + \tfrac{1}{3}r_0(I_p + \Delta I_p) + \frac{n}{6}(V_S + \Delta V_S)\right] - \tfrac{2}{3}V_a. \tag{45}$$

As a numerical example, consider such an internally marked network employing neon gas breakdown switches. For such a device,

$$r_0 = 5000 \text{ ohms,}$$

$$I_p = 10 \pm 3 \text{ milliamperes.}$$

Using the previously given values for the other parameters, the margin M vanishes when $n = 1.5$ stages. This would then appear to be no im-

provement on the end-marked network. For an internally-marked network using devices with smaller values of V_s and I_ρ, full benefit may be taken of the principal advantage of the internally-marked system — the saving in current fanout between successive stages.

From the foregoing discussions, it is possible to arrive at certain conclusions from a switching margin viewpoint concerning desirable switch characteristics. These are listed below.

(a) All margins would be considerably improved if the ac breakdown voltage used in the calculations were equal to the nominal dc value. This would require the switch to take all transient voltages without exhibiting any loss in V_B.

(b) The margins improve as the variations in the switch parameters, ΔV_B and ΔV_s, are reduced. This requires a close supervision of the acceptable limits placed on switch characteristics during manufacture, together with minimum deterioration in the lifetime of their application.

(c) The overvoltage, Δ, required to assure breakdown of the switch either subtracts from the margins or may be considered to be a margin in itself. For gas diodes the speed of switching depends on the overvoltage, and so a compromise must be made between margin and speed. For the p-n-p-n switch the speed limitations due to overvoltage limitations are inconsequential, so that the overvoltage requirement is small.

(d) The power consumed in maintaining a switch in its "on" state depends on the minimum sustaining current, I_0, and the sustaining voltage, V_s. Consequently, a low value of I_0 and V_s will save considerably on the holding power drain.

(e) The forward and reverse impedance of the switch in "off" condition should be as high as possible. Then the impedance at idle nodes consists of the bias resistor connected to the idle bias potential in parallel with the impedance of multipled switches connected to other busy or idle nodes. If the multipled switches terminate at nodes that are not in the idle condition, leakage currents will flow into the node and alter the value of the idle potential. To prevent excessive variation of the idle bias voltage at a node (and hence deteriorate the margins) it is necessary that the "off" switch impedance be very much greater than the bias resistor. Both gas diode and p-n-p-n switches are expected to exhibit adequately high "off" impedance.

X. TRANSMISSION ASPECTS OF A BREAKDOWN CROSSPOINT SWITCHING NETWORK

The busy state of network nodes directly affects the switching margins. This is due to variabilities in the switch sustain potential accumu-

lating from stage to stage and to the amplitude of the speech signals which are transmitted through the paths. These and other interactions between switching and transmission states, as well as transmission factors *per se*, have largely determined the design choices. This fact distinguishes networks of this type from the electromechanical systems, where the switching and transmission factors are separate to a far greater degree and are of relatively minor importance as design limitations.

Discussion of transmission through the gas diode switching network starts with the *V-I* characteristic, Fig. 1. The intersection of the busy-state load line with the switch characteristic sets the operating sustain current at 10 milliamperes. At this current, the ac resistance of the switch is negative. The magnitude of the resistance, about 200 ohms, varies with the current, temperature, age of the switch and, in a way related to the temperature behavior, it varies with time over the first several minutes from the moment the path is made busy.
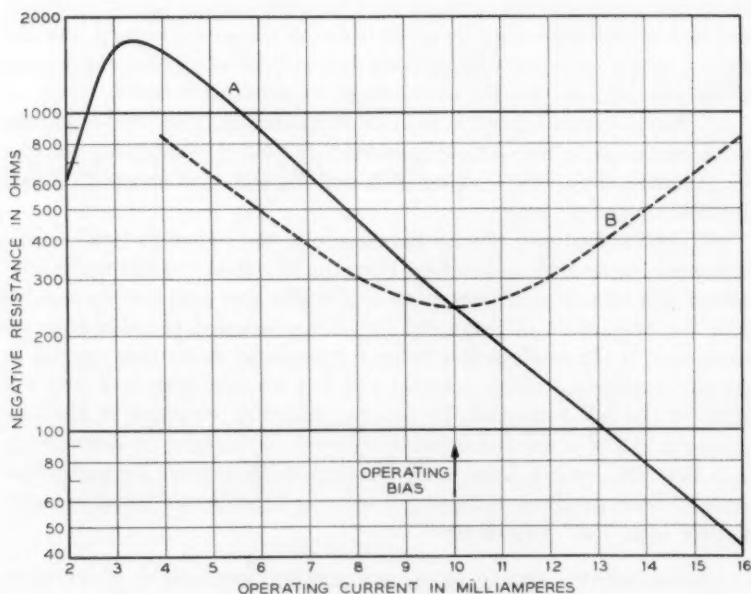

Fig. 7 — Variation of the A1624 gas diode negative resistance with operating current.

Curve A of Fig. 7 illustrates variation of negative resistance with current. The steepness of the characteristic at the operating point affects the design in two ways. First of all, it indicates that the source of holding current must be closely regulated, or variations in the negative resistance will occur from time to time as the power supply varies. These variations produce over-all gain variations which result, in one direction, in high transmission loss through the network and, in the other direction, in oscillations. The switch is useful as a series negative-resistance transmission element for compensation of the incidental series transmission losses elsewhere in the switching network, principally the copper loss of the coupling transformers. Gain compensation in this manner is analogous to the use of negative-impedance two-wire repeaters for trunk circuit loss compensation. The wide variability of the external terminations for the switching network indicates that the network transmission paths should be unconditionally stable between open or short circuit terminations. Excess positive resistance loss is designed into the transmission path to account for the maximum variations in positive and negative resistance and assured network stability. The greater these variations are, the greater must be the residual transmission loss. In an experimental six-stage gas diode switching network, the design value for the transmission loss is 1 db. This value also produces an optimum over-all return-loss characteristic, since the residual shunt and series losses are about equal and together represent a pad properly matched to the 2000-ohm network circuit impedance level.

The dependence of the impedance of the gas diode switch upon current passing through it has a second deleterious effect. Signal current variations produce distortion and, more importantly, if sufficiently large, they swing into the low-current high-gain region of the switch, with consequent instability. Both effects have been minimized in the six-stage network. A potential introduced at the center of the network supplies parallel-feed hold current to each pair of three oppositely poled stages of switching. Signal current flows through all six switching stages in series. Thus, as the signal current increases in three tubes it simultaneously decreases in the remaining three tubes. Compensating negative resistance variations are thereby introduced into the signal path. The resistance of a single tube may be expressed as $R(I) = R_0(I_0) + \Delta r(\Delta i)$, where $R_0(I_0)$ is the resistance at the bias current I_0, and $\Delta r(\Delta i)$ expresses the change in resistance produced by deviations in current from the bias value. The resistance of six tubes in a center-fed path is given by:

$$\sum_6 = 6R_0(I_0) + 3\Delta r(\Delta i) + 3\Delta r(-\Delta i)$$

$$= 6[R_0 + \tfrac{1}{2}\Delta r(\Delta i) + \tfrac{1}{2}\Delta r(-\Delta i)].$$

To the extent that tubes are alike, an even-order balance is achieved. The effective resistance per tube is plotted as curve B of Fig. 7, which is a construction from curve A assuming a mean bias of 10 milliamperes and signal excursions about this operating point. The characteristic, parabolic in shape, results in an over-all increase in gain at the peaks of signal excursions and forces operating range limits of about ±3.5 milliamperes from a 10-milliampere bias. These current limits at the 2000-ohm network transmission impedance permit a maximum sine wave transmission amplitude of 12 milliwatts. The current limiting circuit in the junctor is biased to limit the current excursions to the ±3.5-milliampere value. This current swing produces a corresponding maximum signal voltage swing of ±7 volts at the terminals. The voltage clipper diodes in shunt with the transformer winding are designed to conduct at ±9 volts, just above the maximum signal. These diodes limit input signal amplitudes to values which can be transmitted. Their basic purpose is protection. In combination with the current-limiting feature, they prevent lightning and power faults from breaking down switches into existing network connections and from extinguishing paths which are present. Without this protection, early networks were sensitive to excessive signals; the paths were easily extinguished by extreme speech signal peaks.

Another significant characteristic of the gas diode switch is its transmission behavior as a function of temperature. Ambient temperature acts directly to change the negative resistance. In addition, temperature variations produce changes in the sustain voltage of the switch. These voltages, in turn, react to produce changes in the bias current of the held path, with resulting resistance changes already discussed. In addition to the effects of ambient temperature, the tube when operated dissipates about 1 watt, to produce a relatively large temperature rise in the first several minutes of operation. Both temperature effects were overcome by at all times maintaining about the switches a stream of air held at fixed temperature (±5 Fahrenheit degrees) and moving about 500 feet per minute. By this expedient the temperature of the tube is held within narrow limits. Fig. 8 shows the variation in the over-all six-stage network input resistance with time, with and without the benefit from moving air, from the moment of establishing a path. Curve B illustrates an increasing impedance with time corresponding to a decrease in the
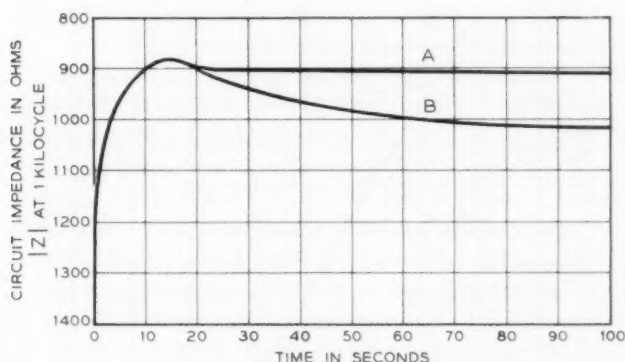
Fig. 8 — Starting circuit impedance transient due to the thermal properties of the A1624 negative-resistance characteristic: (a) air flow of 500 feet per minute across tubes; (b) tubes in stagnant air.

negative resistance of the gas diode switches as they heat in still air. Curve A corresponds with curve B but, in this case, with the use of a moving air environment. The expedient taken in this case to overcome the temperature characteristic is costly and was employed in view of the short time available. A promising longer range approach includes a better gas discharge characteristic and better packaging from a thermal point of view.

XI. CROSSTALK

The 2000-ohm network impedance level for the gas diode switching network is a compromise choice between insertion loss and capacity-coupled crosstalk among busy paths. Coupling among the busy paths is a special problem, since the switching network is unbalanced to ground. The ground plane within the equipment provides a common return circuit for all the path-terminating transformers. The coupling impedance of the common ground connection has not been a serious limitation. Mutual capacitance between adjacent circuits is the limiting source of coupling. It is minimized by switch designs which attempt to secure a maximum separation between adjacent circuits. The links between stages of switching necessarily are several feet in length and their large number necessitates relatively close spacing. In the present
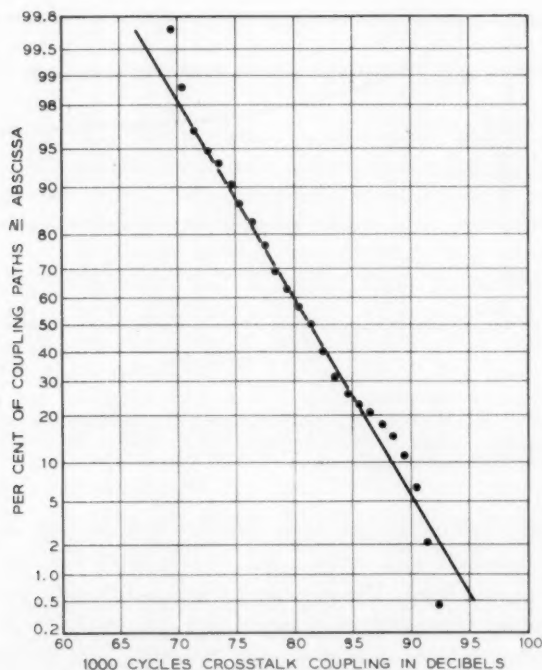
Fig. 9 — Distribution of crosstalk loss values for an experimental six-stage gas diode switching network.

design, direct coupling was reduced by the shielding effect of ground wires interspersed with the active leads. Factors of improvement of 3 to 10 are readily obtained with no more shielding than this.

At the 2000-ohm circuit impedance level, the mean path-to-path crosstalk loss achieved in the present design is about 80 db. Fig. 9 shows the distribution among paths to be nearly normal, with a sigma of 5 db. This performance is considered to be just tolerable and an improvement of 5 db is to be sought in subsequent designs. Higher circuit impedance would further aggravate the crosstalk problem and lower impedances will be beneficial. However, from the standpoint of insertion loss variations due principally to gas diode negative resistance variations, the situation is reversed. Higher circuit impedances will better absorb the switch impedance variations and lower impedance circuits would be more susceptible. For the experimental gas diode switching network, the impedance was set at 2000 ohms, as high a value as seems possible in terms of crosstalk performance.

XII. ACKNOWLEDGMENTS

The development of a multistage switching network is the result of the cooperation of many persons. The authors wish to express their appreciation for the contributions to this project by their many co-workers. In particular we wish to thank R. J. Andrews, R. Biddulph, K. S. Dunlap, L. W. Hussey, G. E. Jacoby, W. Peil, R. L. Simms and E. A. Woodin for their circuit contributions. Although the mechanical aspects of the experimental electronic switching system have not been discussed in this paper, the system itself would not have been possible without the mechanical design considerations of R. A. Billhardt, W. Hopf, J. H. Mogler, R. B. Mutke and L. J. Trimnell. The authors are also indebted to R. W. Ketchledge for his guidance and many valuable suggestions.

REFERENCES

1. Joel, A. E., An Experimental Switching System Using New Electronic Techniques, B.S.T.J., **37,** September 1958, pp. 1091–1124.
2. Lee, C. Y., Analysis of Switching Networks, B.S.T.J., **34,** November 1955, pp. 1287–1315.
3. White, A. D., A Novel Form of Hollow Cathode and Its Discharge Characteristic, 9th Annual Gaseous Electronics Conference, November 1956.
4. Brewer, S. T., and Hecht, G., A Telephone Switching Network and Its Electronic Controls, B.S.T.J., **34,** March 1955, pp. 361–402.

The Timing of High-Speed Regenerative Repeaters

By O. E. DE LANGE

A simplified method for determining the behavior of the timing portion of a chain of regenerative repeaters is presented. The method applies to self-timed as well as separately timed systems. The effect of random noise, as modified by the timing circuit bandwidth, is calculated. Defects other than noise are discussed. System requirements for the satisfactory transmission of a number of types of signals are determined.

I. SCOPE OF PAPER

Technical advances in the microwave art, and particularly in connection with waveguide transmission, have made pulse code modulation with the use of regenerative repeaters a plausible means for broadband transmission over distances of thousands of miles. Such transmission can be successful only if the signal can be satisfactorily retimed in a chain of one to a few hundred repeaters.

The problem of retiming has been considered, chiefly in connection with a particular type of repeater, by E. D. Sunde.[1] Extensive mathematical treatments of this problem appear in accompanying articles by W. R. Bennett[2] and H. E. Rowe[3] of Bell Telephone Laboratories. This paper points out some of the problems encountered in timing and presents a less mathematical method for determining the behavior of the timing portion of a chain of regenerative repeaters. It also considers the behavior in connection with the requirements for the satisfactory transmission of broadband signals through 100 repeaters.

The method used here is based on the fact that the behavior of the timing part of a chain of self-timed binary repeaters is very similar to that of a chain for which timing information is transmitted over a separate analog channel. Both of these systems, therefore, can be analyzed by methods which are known for analog transmission. The method is applied to determine the effects of random noise upon a single re-

1455

peater and it is then extended to include a chain of repeaters. The effect of timing circuit bandwidth is determined. The method yields numerical answers which agree with experimental results, as pointed out in the following article.[4] It is found that, although noise effects accumulate along a chain of repeaters, the rate of accumulation is low.

Effects other than those due to noise are discussed. In a realistic case these effects appear to be more severe than those of noise. Since they result from equipment deficiencies, it should be possible to reduce them to tolerable values by employing improved timing circuits. In particular, this can be accomplished by means of a few slightly more complicated timing circuits distributed throughout a system employing simple tuned circuits as timing filters.

We conclude that, although noise and effects resulting from changes of pulse pattern preclude the possibility of regenerating PCM signals an indefinite number of times, it is possible to regenerate the few hundred times required for a long system without having to resort to complicated timing equipment or provide a signal-to-noise ratio much greater than that imposed by system requirements other than those involved in timing.

II. GENERAL CONSIDERATIONS

The main advantage of transmitting information in binary-pulse form results from the fact that such pulses can be regenerated at each repeater (see Ref. 5). The process of regeneration removes most of the signal distortion produced by noise, bandwidth limitations, etc., during transmission. If the distortion is not allowed to become excessive between regenerators the signal at the end of a system with many repeaters might have little more distortion than the signal at the input to the chain of repeaters.

The regeneration of binary pulses involves two functions: (1) the removal of undesirable amplitude effects; (2) the restoration of each pulse to its assigned position in time. An ideal repeater provides a perfect train of timing pulses to perform the second of these functions. In a practical repeater, these pulses may be displaced from their correct positions as a result of noise or other disturbances. The fundamental timing problem is that of deriving suitable timing pulses in the presence of such disturbances. There are two basic methods of providing timing information at a repeater: in the first method, this information is transmitted over a separate channel, and the system is said to be "externally timed"; in the second, it is derived from the incoming train of signal pulses to provide a "self-timed" system.

In either of the above systems the timing wave will be recovered as an approximately sinusoidal wave through a narrow-band filter of some kind. This sine wave is converted to a train of short pulses, the times of occurrence of which are determined by the phase of the sine wave. Although the timing function is performed only by the train of pulses, we shall use the term "timing wave" to indicate either this train or the sine wave from which it is derived. Along with the sine wave at the output of the filter there will be undesired frequency components due to any disturbances which may exist. The resultant wave is caused to vary in phase and amplitude as shown for a single interfering component by Fig. 1(a). Limiting can be employed to remove the amplitude variations but the phase deviations, as shown by Fig. 1(b), are of the most concern in a timing system. The amount of phase deviation is determined by the ratio of timing signal power to timing noise power. We shall, therefore, direct a considerable amount of attention to the calculation of this ratio.

When the disturbing power entering the timing circuit consists of several components or of a large number spread over some frequency range, the resultant effect is equal to the sum of the effects of the individual components if we limit consideration to the case where the
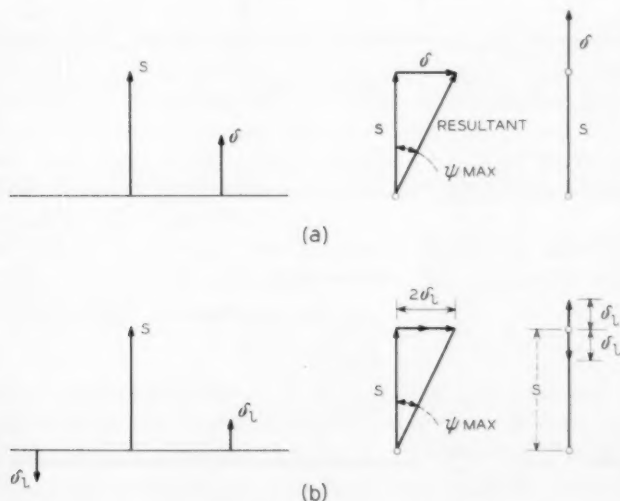

Fig. 1 — Effect of a single interfering frequency (a) before limiting and (b) after limiting.

signal-to-noise ratio is large. It is obvious that, for the distributed spectrum, the narrower the band of the timing filter, the fewer the number of components it passes and the smaller the phase deviation will be. From a practical standpoint, stability considerations place a limit on how narrow this band can be made; i.e., the narrower the band of the circuit, the more subject it becomes to slight detunings brought about by temperature variations, etc.

III. EXTERNALLY TIMED SYSTEMS

3.1 *Straight-Through Timing Wave Repeater*

Fig. 2(a) shows an externally timed system with one timing channel supplying two signal channels. Fig. 2(b) shows, in greater detail, a
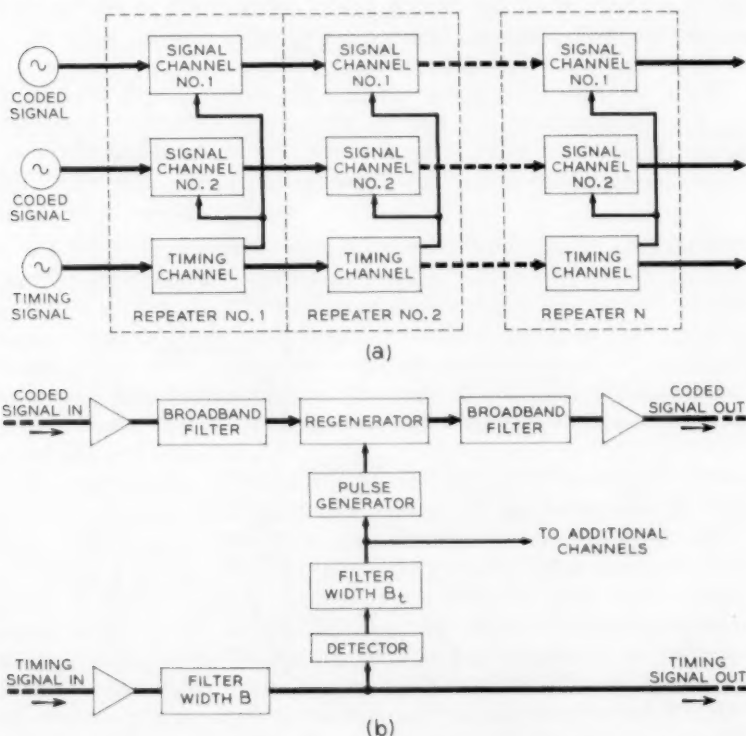
Fig. 2 — (a) Externally timed system with one timing channel supplying two signal channels; (b) single repeater—straight-through timing channel.

single repeater of such a system. The straight-through timing channel shown on Fig. 2(b) appears to provide the simplest form of timing circuit. Here a sine wave at pulse-repetition frequency f_r is amplitude modulated* on an RF carrier of frequency f_c and transmitted straight through the system. At each repeater this RF signal is simply amplified, filtered and sent on to the next repeater. Each RF circuit, including filters, is assumed to have a frequency characteristic which is flat over a total band of width $B = 2 f_r$ so as to pass the amplitude modulation sidebands. The characteristic is assumed to cut off sharply outside of this range. It is assumed that all noise in the system originates at the input to each repeater, that the RF signal-to-noise ratio is always large and that there is unity gain from repeater input to repeater input. The timing channel of Fig. 2(b) is seen to be a straightforward analog system in which, at the end of N repeaters, the noise power will be N times that at a single repeater. The RF signal-to-noise ratio at any repeater is given by

$$\left(\frac{W_s}{W_n}\right)_{\mathrm{RF}} = \frac{P_s}{N W_0 B},$$ (1)

where P_s is the mean RF power at the peak of an envelope cycle† and W_0 is the noise power density, in watts per cycle of bandwidth, contributed by a single repeater.

At each repeater some signal is taken off through a branch circuit where it is detected to recover the envelope, which is filtered through the narrow-band timing filter of bandwidth B_t to become the timing wave at that repeater. Although noise power is available over the entire band B, those components lying within two bands of width B_t and spaced by frequency f_r on either side of the RF carrier contribute most of the noise to the timing circuit.

From (23) of Appendix A we find the timing wave signal-to-noise ratio at the output of the narrow filter of the first repeater to be

$$\frac{W_s}{W_{na}} = \frac{P_s}{8 W_0 B_t},$$ (2)

where W_s is the mean power of the sinusoidal timing wave and W_{na} is the mean noise power. From a consideration of the way noise accumulates in a system, it is evident that, for N repeaters, the timing signal-

---

* Although it should be possible to transmit timing information by other means, for example by frequency modulation, we shall limit consideration in this paper to amplitude modulation systems.

† Power is specified in this manner rather than in the more usual terms of RF carrier power since, in later discussions of pulses, peak power is the significant quantity.

to-noise ratio is

$$\left(\frac{W_s}{W_{na}}\right)_N = \frac{P_s}{8NW_0B_t}.\tag{3}$$

3.2 Remodulation Timing Wave Repeaters

Fig. 3 shows a single repeater for another type of separate timing system. Here, at each repeater, the envelope of the RF wave is recovered by the detector, filtered through the narrow filter of width B_t and remodulated onto a locally generated RF wave for transmission to the next repeater. The timing signal-to-noise ratio at the output of the narrow-band filter of a remodulation repeater will be the same as that at the corresponding point in a straight-through repeater if both employ ideal, flat timing filters. This ratio for a chain of N repeaters is, therefore, given by (3).

If instead of using a local oscillator, the incoming signal is employed as a source of RF and, if the modulator is replaced by a time gate, the repeater of Fig. 4 evolves from that of Fig. 3. This configuration does not require a local oscillator but has the disadvantage that noise can be passed through the gate via a second path which does not include the timing filter. This difficulty can be overcome by replacing the gate with an ideal pulse regenerator. Such a regenerator produces an output which consists entirely of off-on pulses, the times of occurrence of which depend only upon the timing wave. Thus, no timing noise is passed through the second path except in the form of errors in the pulse pattern. These errors have negligible effect upon timing.

If the gate is replaced by a regenerator, the result is the repeater shown on Fig. 5. It is evident that this repeater can also serve as a self-timed repeater in a signal channel. We thus arrive at the same configura-
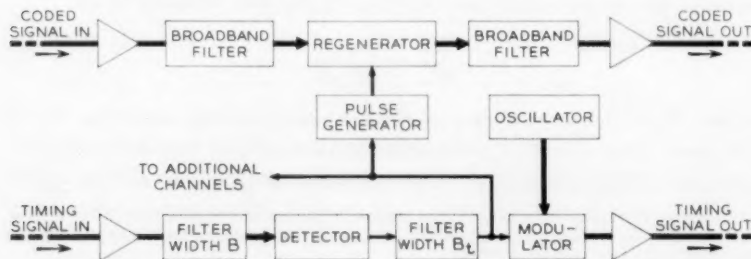

Fig. 3 — Single repeater—remodulation timing channel.

tion for a self-timed, signal-channel repeater and a desirable form of repeater for a timing channel. If we adopt this as a preferred form of timing-wave repeater (as we shall in this paper), the most significant difference between a timing channel and a chain of repeaters carrying a complete signal lies in the type of signal transmitted. For the timing channel, the signal envelope will be a sine wave at pulse-repetition frequency at most points in the system; for the complete channel, the signal envelope will consist of the varying pattern of pulses which results from coding of the information being transmitted. The timing portion of the repeater shown on Fig. 5 is seen to consist of a detector, a narrow-band timing filter, an amplifier, a limiter and a pulse generator. When the repeater is being used as part of a separate timing channel the timing wave for the signal channels can be taken off at the output of the limiter.

For the systems of Fig. 4 and Fig. 5 the signal envelope exists in the form of pulses at some points in the repeater; however, if the output of
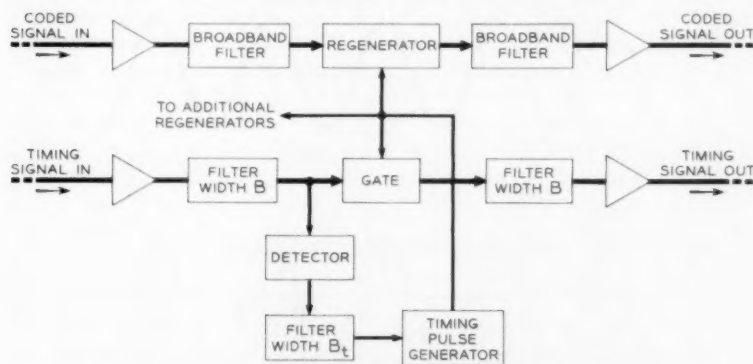

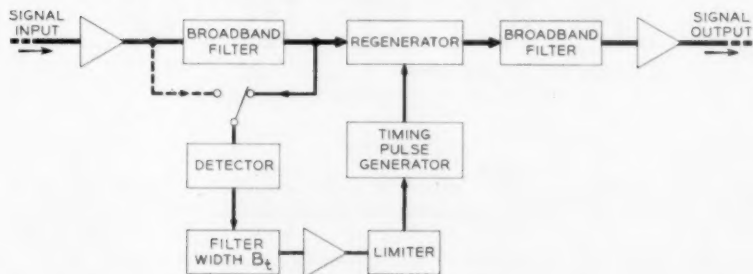Fig. 4 — Single repeater—remodulation timing channel.

Fig. 5 — Self-timed repeater.

the gate or regenerator is passed through a suitable RF filter the envelope is reduced to a sine wave of frequency f_r when all pulses are present. In all cases involving an externally timed system, the signal input to the detector can thus be made to consist of an RF carrier 100 per cent amplitude modulated by a sine wave at pulse-repetition frequency, f_r [see Fig. 6(b)]. This is also true for a self-timed repeater when all pulses are present. In a later section of this paper the effect of RF bandwidth is discussed in more detail.

In comparing the various systems we see that, for the straight-through timing system of Fig. 2, noise is passed from repeater to repeater unchanged in form and builds up over the entire RF band B, as shown on Fig. 7(a). For the remodulation system of Fig. 3 the noise existing over a narrow band of width B_t is passed along a chain of repeaters by being amplitude-modulated on the RF carrier along with the timing wave. The resultant RF sidebands occupy a narrow band above the carrier

$\frac{A}{2} \cos 2\pi f_r t$ (a)

$\frac{A}{2} (1 + \cos 2\pi f_r t) \cos 2\pi f_c t$ (b)

0.5A

0.25A   0.25A
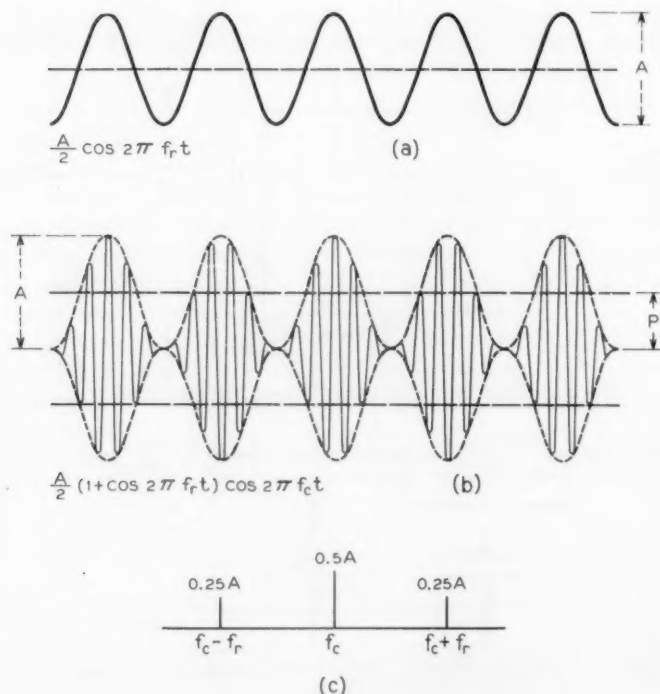
$f_c - f_r$ f_c $f_c + f_r$

(c)

Fig. 6 — Timing signal: (a) envelope; (b) RF signal; (c) frequency spectrum.

and an equal band below the carrier, and build up as indicated in Fig. 7(b). The sidebands above the carrier are correlated with corresponding ones below the carrier, since they originate from the same noise source. For the system of Fig. 5, noise components existing throughout the narrow band B_t time-modulate the envelope of the signal out of a repeater. After filtering, the noise output of one of these repeaters is the same, from a qualitative standpoint, as the noise out of a remodulation repeater [see Fig. 7(b)].*
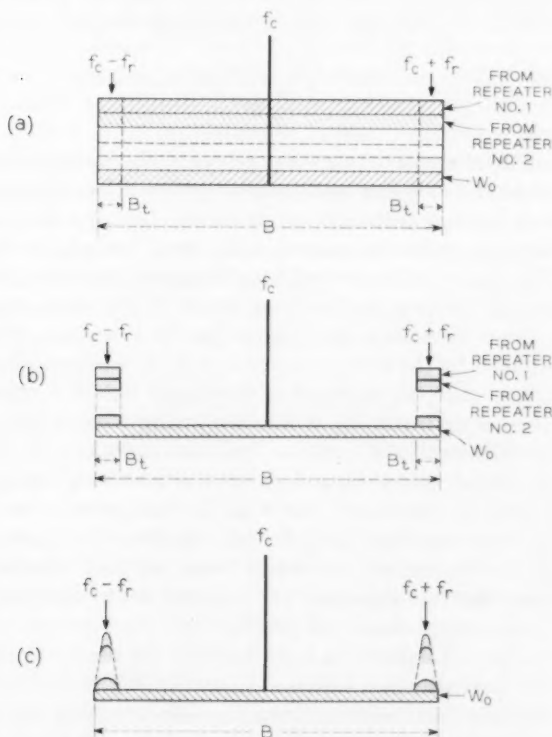
Fig. 7 — Buildup of noise in a chain of repeaters: (a) straight-through repeaters; (b) remodulation repeaters with flat filters; (c) remodulation repeaters with tuned circuits.

* This figure is somewhat idealized in that it assumes a clean, noise-free carrier at each repeater. In an actual system the carrier will be accompanied by noise components, so that at the end of the system there will be some noise power distributed throughout the band B.

3.3 S/N Ratio in the Presence of Random Noise — Flat Filters

At the output of the timing filter of the first repeater of the type shown on Fig. 5 the signal-to-noise ratio is the same as for the straight-through repeater or the remodulation repeater and is given by (2):

$$\frac{W_s}{W_{na}} = \frac{P_s}{8W_0B_t}. \tag{4}$$

In the process of limiting, the components of noise in phase with the timing sine wave are removed so that the signal-to-noise ratio becomes

$$\left(\frac{W_s}{W_n}\right)_1 = \frac{P_s}{4W_0B_t}. \tag{5}$$

This is the ratio of mean timing wave power to the mean power carried by the phase modulation sidebands resulting from noise. The mechanism of phase modulation and the sidebands which exist after the signal and a single-frequency noise component have been limited are shown by Fig. 1(b). The timing pulses derived from the phase-modulated sinusoidal timing wave will deviate in time as a result of this noise modulation. Since the time of occurrence of a pulse out of the time gate, or regenerator, is determined by the time of occurrence of the corresponding timing pulse, it follows that the envelope of the signal out of a repeater will deviate in time in the same way as does the timing wave in that repeater.

If the second repeater of a chain were free of local noise its timing wave would follow phase excursions of the incoming signal and be phase-modulated to the same extent as the timing wave at the first repeater. (If both repeaters have flat timing filters the noise sideband pattern will be transmitted unaltered from the first repeater to the second.) Since the two repeaters are assumed to be identical, locally generated noise acting alone will produce the same average phase deviation in the second repeater as is produced in the same way in the first repeater. The average noise sideband power arising from local noise is therefore equal to that resulting from time deviations of the incoming signal. Since there is no correlation between the sidebands resulting from these two sources of noise they add on a random, or power, basis. For flat filters the timing signal-to-noise ratio at the second repeater is therefore one-half that at the first repeater and, at the Nth repeater, is

$$\left(\frac{W_s}{W_n}\right)_N = \frac{P_s}{4NW_0B_t}. \tag{6}$$

3.4 *S/N Ratio in the Presence of Random Noise — Tuned Circuits*

For the system of flat filters discussed above, the bandwidth of the total timing system is the same as the bandwidth of a single filter. With tuned circuits, or other peaked filters, the system bandwidth decreases continuously as the number of repeaters (and hence the number of filters) is increased. With tuned timing filters, noise builds up as shown by Fig. 7(c). At the first repeater the noise input to the detector is the same as for the repeater with a flat filter. The input to the detector at the second repeater differs from the corresponding input for a repeater with a flat filter in that the components from the first repeater have a different frequency distribution. If we now consider the second repeater to be free of locally generated noise, it is no longer true that the timing wave out of its filter is deviated to the same extent as the wave at the corresponding point in the first repeater. In this case, some of the sidebands about the timing wave suffer attenuation in the second filter and similarly at succeeding repeaters. The noise power which originates in the first repeater is filtered by N circuits before it reaches the end of the system; that which originates in the second repeater is filter by $(N - 1)$ circuits and so on, as shown on Fig. 7(c).

In Appendix B it is shown that, for a system of N repeaters of this type, the timing signal to-noise ratio becomes

$$\left(\frac{W_s}{W_n}\right)_N = \frac{P_s Q}{7.1 \sqrt{N} W_0 f_r}, \tag{7}$$

where Q is the "quality factor" of the tuned circuits and the other quantitites are as specified previously. Equation (7) indicates that, with tuned circuits, noise power increases as the square root of the number of repeaters. For flat timing filters the noise power increases directly as the number of repeaters, as shown by (6). This advantage of tuned circuits over flat filters seems to have been first pointed out by H. E. Rowe.[3]

3.5 *Timing Displacement*

Having obtained the timing wave signal-to-noise ratio for a chain of repeaters, it is fairly simple to determine the rms magnitude of the phase deviations of the timing wave. First, consider the noise to be replaced by a single-frequency interference having the same mean power. From Fig. 1(b) it is evident that the following approximate relationships hold:

$$\Psi_{max} = \frac{2\delta_l}{s} \quad \text{and} \quad \Psi_{rms} = \frac{\sqrt{2}\,\delta_l}{s},$$

where δ_l is the magnitude of one of the sidebands of the amplitude limited timing signal and s is the amplitude of the sine wave. Returning to the equivalent noise we have

$$\frac{\sqrt{2}\,\delta_l}{s} = \sqrt{\frac{W_n}{W_s}}, \quad \text{or} \quad \Psi_{rms} = \sqrt{\frac{W_n}{W_s}}. \tag{8}$$

Using the value of signal-to-noise ratio given by (7) for a chain of repeaters with tuned circuits and the above expression for Ψ_{rms} yields

$$\Psi_{rms} = 2.66 N^{1/4} \sqrt{\frac{W_0 f_r}{P_s Q}}. \tag{9}$$

It should be noted that, although the rms phase deviation produced by noise is the same as the rms deviation produced by the equivalent single frequency, the peak deviations will be different for the two types of interference. Therefore to determine the peak phase deviation the appropriate peak factor must be applied to (9).

IV. SELF-TIMED SYSTEMS

Since we have chosen the same repeater configuration for externally timed and self-timed systems, the main difference between the two lies in the types of signals each transmits. When a chain of repeaters is used as a separate timing channel for one or more signal channels it transmits a fixed pulse pattern or a sinusoidal timing wave, whereas when used as a self-timed system the chain transmits the varying pulse pattern which results from coding a baseband signal. For a self-timed system there are gaps in the train of signal pulses applied to the timing circuit filter. Improved timing signal-to-noise ratio is obtained if the circuit is so arranged that no noise is applied to the filter during these gaps. For a baseband system this can be accomplished by inserting a nonlinear device, such as a peak amplifier, ahead of the filter and adjusting it so that only voltages exceeding some threshold value at the input will be passed to the filter. This is discussed in greater detail in the following article. If, in a carrier system, a square-law detector is employed to recover the pulse envelopes which supply the input to the timing circuit, this detector can be made to perform much the same function as the nonlinear device mentioned above. This is pointed out by Rowe.[3] Because of its advantage over the linear detector in this respect, the square-law detector is the only type which will be considered here. For repetitive patterns in which only one pulse position out of M possible positions is occupied, it follows from a consideration of the Fourier com-

ponents involved that the component at pulse-repetition frequency, f_r, is reduced in amplitude by the factor $1/M$. The more general expression for timing wave power then becomes, from (20) and (21) of Appendix A:

$$W_s = \frac{\alpha^2 A^4}{32 M^2},\tag{10}$$

where α is the detection coefficient of the particular detector used and A is the peak RF amplitude.

If we assume that, because of the nonlinearity of the detector, the noise power applied to the timing filter is proportional to the number of pulses per unit time, the more general expression* for noise power becomes, from (22):

$$W_{na} = \frac{\alpha^2 A^2 W_0 B_t}{2M}.\tag{11}$$

Since it evaluates the noise in a single repeater, (11) applies to either flat timing filters or tuned circuits, as long as B_t is taken as the effective bandwidth.

From (10) and (11) we obtain the timing wave signal-to-noise ratio for a single repeater at the output of the timing filter, before amplitude limiting:

$$\left(\frac{W_s}{W_{na}}\right)_1 = \frac{A^2}{16 W_0 B_t M} = \frac{P_s}{8 W_0 B_t M}.\tag{12}$$

After limiting, it is

$$\left(\frac{W_s}{W_n}\right)_1 = \frac{P_s}{4 W_0 B_t M}.\tag{13}$$

Equation (13) differs from (5) only by the factor M in the denominator of (13). Similarly, from (7), the signal-to-noise ratio for a chain of repeaters with tuned timing filters becomes

$$\left(\frac{W_s}{W_n}\right)_N = \frac{P_s Q}{7.1 \sqrt{N}\, M W_0 f_r},\tag{14}$$

and

$$\Psi_{rms} = 2.66 N^{1/4} \sqrt{\frac{M W_0 f_r}{P_s Q}}.\tag{15}$$

---

* Although this expression is only approximate, the approximation has been found to be sufficiently good from a practical standpoint.

4.1 *The Effects of Variations in Pulse Pattern*

The equations derived previously are based on repetitive pulse patterns, so that caution must be observed in applying them to a system for which the pulse pattern is changing. If we assume that the probability of a pulse being present in any pulse position is equal to one-half, the M of (15) has a long-time average value of 2. However, if no restrictions are placed on the pattern of pulses to be transmitted, there may be comparatively long periods during which there are no pulses present. Unless the timing filter has an infinite Q, the timing wave may decay to an unusable value during these periods. Fortunately, it is possible, with only a small sacrifice of information-handling capacity, to code messages in such a way as to guarantee that the longest gap will be only a few tens of pulse intervals. This is accomplished by omitting the one amplitude level corresponding to no pulses in the code group. Even with this type of coding it is possible with some signals, such as television, to have relatively long periods when considerably fewer than one-half of the pulses are present. During these periods M will have a value greater than 2 and the phase deviations will be correspondingly greater. Some other effects of changing pulse pattern will be discussed in a later section.

4.2 *The Effects of Radio-Frequency Bandwidth*

All of the analysis leading to (15) was based on consideration of a radio-frequency band of sufficient width to resolve adjacent pulses and thereby produce the wave of Fig. 6(b) when all pulses are present. Both theory and experiment indicate that a system should provide satisfactory amplitude regeneration with considerably less than this amount of bandwidth. Since conservation of bandwidth is always desirable, we proceed to study the effects upon timing of reducing the bandwidth to something less than that required to resolve adjacent pulses. Frequency distributions and a possible filtering arrangement for adjacent PCM channels are shown in Fig. 8, where two adjacent carriers are represented by f_{c1} and f_{c2}. The transmitting filter in each channel corresponds to the filter following the regenerator in Fig. 5. Returning to Fig. 8, we see that the transmitting filters have some loss to the sidebands spaced by the repetition frequency from the carriers. This loss is necessary to obtain the wave of Fig. 6(b) from the gated output and it can become as great as 6 db if the gating pulse is extremely short. The channels are shown separated by a guard band B_g. If components from one channel fall beyond this guard band they may interfere with an adjacent channel,

and especially with its timing circuit. Furthermore, once these components enter the common medium, filtering is not effective against them, since the desired components would be reduced to very nearly the same extent as the undesired. This is evident from the receiving filter characteristic of Fig. 8 when one considers that all usable timing information, as well as timing disturbances, is contained in the bands of width B_t spaced either side of a carrier by the frequency f_r.

If the timing detector is connected as shown by the solid line in Fig. 5 the receiving filter, which follows the input amplifier, will attenuate timing signal and timing noise occurring in the narrow bands of width B_t to very nearly the same extent, as can be seen from Fig. 8. Then, from the standpoint of the timing signal-to-noise ratio, it makes little difference what characteristic this filter has, though some disadvantage would follow from too large a reduction in the absolute value of the timing wave. It should, therefore, be advantageous to connect the timing-wave
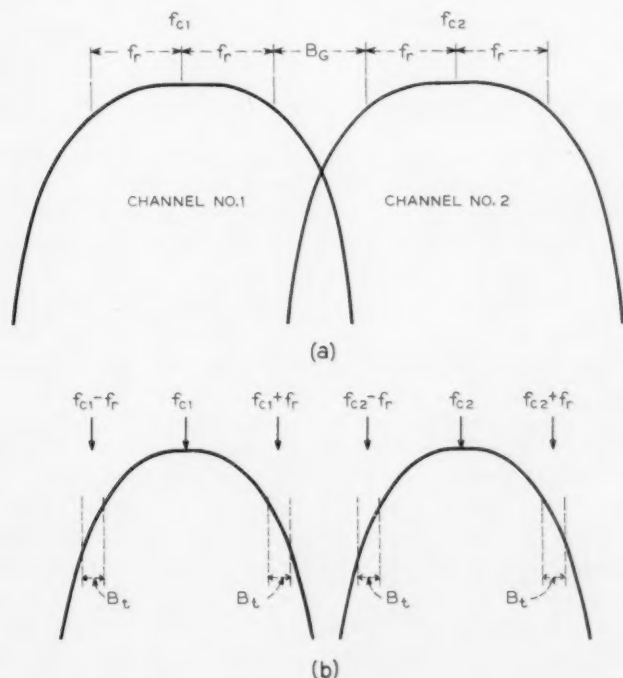

Fig. 8 — Filter characteristics: (a) transmitting filter; (b) receiving filter.

detector ahead of the filter, as shown by the dashed line of Fig. 5, in order to avoid timing as a consideration in the design of this RF filter. (For broadband RF amplifiers it would be desirable to retain some filtering ahead of the detector in order to maintain a large signal-to-noise ratio at the detector input.)

We conclude that, from the standpoint of timing, the transmitting RF filter at each repeater is the important one. For such filters with characteristics matched to the length of pulse out of the regenerator so as to produce the wave of Fig. 6(b), the results expressed by (15) apply. If the transmitting filter is made narrower than shown so as to have some additional loss* at frequencies spaced from the carrier by f_r, there will obviously be a reduction in the timing signal-to-noise ratio. Proper slicer operation can be obtained for bands considerably narrower than that required for resolution of adjacent pulses. It is therefore possible to make the receiving filter ahead of the regenerator considerably narrower, as indicated by the characteristic at the bottom of Fig. 8. Such filtering provides additional discrimination against noise and other interference which might affect the slicer.

4.3 *Effects Other Than Noise*

Noise is not the only source of timing difficulty in a self-timed system. Assume, for example, a defect in the timing circuit such that the phase of the timing wave which it delivers is a function of the pattern of pulses being transmitted. Then any change of pulse pattern may result in a change of timing wave phase. Such defects exist in the form of amplitude-to-phase conversion in limiters and other nonlinear elements, phase changes due to the finite width of the pulses applied to the timing circuit and phase shifts due to detuning of the timing circuits. Although the results are somewhat similar to those produced by noise they differ in two important respects: (1) Except for detuning, it is very probable that the deviation will be the same at each repeater, so that deviations add directly rather than on a random basis. It is also possible, though highly improbable, for deviations due to detuning to add directly. (2) These defects are not inherent in timing circuits and might conceivably be reduced to a negligible value by proper design of timing equipment.

For a chain of repeaters having defects of this type, transmission

---

* It appears to be possible to devise systems which provide satisfactory timing even when the band is limited to such an extent that this loss becomes considable. Some extreme cases are discussed by Bennett.[3] Such systems do not lend themselves to simple analysis and therefore will not be dealt with here.

delay is a function of the pulse pattern, which, in turn, is a complicated function of the signal being transmitted. This is somewhat analogous to delay distortion in an analog system. The effect of changes of pulse pattern is not amenable to general treatment, since it is a function of the particular timing circuit being considered. The defect has been evaluated by Sunde[1] for a model based on one particular type of repeater. Our experience up to the present time leads us to believe that, unless some specific remedy is applied, this problem may be more serious than that posed by random noise. The frequency spectrum of the phase deviations resulting from changes of pulse pattern is determined by the rates at which the pattern changes and the bandwidth of the timing circuit in which the deviations are produced. The spectrum at the output of any repeater will be further modified by all succeeding repeaters. The solution to the changing pulse pattern problem therefore appears to be that of providing narrow-band high-stability filters in the timing circuits.

4.4 Detuning

Although detuning has already been considered along with the other phase-conversion effects, it warrants further consideration. All calculations up to this point have been based on the assumption that the timing circuits are always perfectly tuned to the pulse-repetition frequency, f_r. This requirement is difficult, if not impossible, to fulfill; we therefore attempt to determine the effect on timing of small amounts of detuning. We assume that each circuit was originally perfectly tuned and that delays were then adjusted to provide perfect alignment between signal pulses and timing pulses. Any subsequent change of tuning will result in a change of phase delay through the circuit and consequently a departure from alignment between signal and timing pulses. An excessive amount of alignment displacement can result in errors in pulse pattern. Fortunately, the steady-state phase displacements do not accumulate in a chain of repeaters except as a change of total delay through the system. Such changes of system delay do not degrade system performance, as long as they take place slowly.

W. R. Bennett has shown[2] that, in addition to the steady displacement discussed above, there is a dynamic displacement of timing-wave phase when the pulse pattern varies with time; i.e., there are additional variations of phase about the displaced position. When the pulse pattern varies in a random manner such that the probability of a pulse being present at any pulse position is one-half, the resultant phase fluctuations

are very similar to those produced by random noise. He has determined that, under these conditions,

$$\epsilon = K\sqrt{\pi Q}, \tag{16}$$

where ϵ is the rms phase deviations resulting from detuning at a single repeater and K is the tuning error in cycles per second divided by the pulse-repetition frequency.

The manner in which dynamic phase deviations build up along a chain of repeaters is governed by the way the circuits are detuned. If the detunings are random, the phase deviation increases slowly as the number of repeaters is increased; however, if all of the circuits are detuned in the same way the phase deviations add directly from repeater to repeater.

Although there is some uncertainty as to how the dynamic deviations due to detuning build up along a repeater chain, we can determine approximately the maximum amount of detuning which can be tolerated under the most unfavorable conditions and still meet the most stringent system requirements. Although the calculations are based on some approximations regarding the spectral distribution of the phase modulation sidebands, the results are probably correct to within a factor of two or better and should prove useful in determining a bound on tolerable detuning. Assume that all circuits are detuned in exactly the same way and that the system must meet voice multiplex requirements as calculated in Section V. Then, for a chain of 100 repeaters with timing circuits having a Q of 100 at 160 mc, the maximum tolerable amount of detuning can be calculated from (16) to be about three parts in 10^4 if dynamic phase deviations are to be kept to the allowable value.* If the timing circuits in the various repeaters are detuned in a random manner — and this is the most probable condition — the amount of detuning which can be tolerated should be considerably greater than three parts in 10^4.

The steady-state alignment displacement, β, occurring at a single repeater can be expressed approximately as $\beta = 2QK$, where β is in radians and K is the tuning error as before. If K is made equal to 3×10^{-4} to meet the dynamic requirements as determined above we calculate a value of 0.06 radian for β. This is reasonably small, so that, under the specified conditions, the tuning requirements are determined by the dynamic effects; under different conditions, the requirements on tuning

---

* Although these deviations add in phase from repeater to repeater, the total deviation for 100 repeaters is not 100 times that for a single repeater because of the decreased bandwidth of the chain. By a method similar to that of Appendix B, it can be shown that the deviation is about 30 times that for a single repeater if we assume a flat spectrum.

accuracy may be set by limitations on static misalignment. If, to determine the limitations imposed by static displacement, we take 0.2 radian as the maximum allowable value of β and assume a circuit Q of 100, we calculate a value for K of 1×10^{-3}. We conclude that timing circuits with a Q of 100 should never be allowed to become detuned by as much as one part in 10^3, but that detuning by one part in 10^4 would be tolerable even under the most unfavorable conditions.

4.5 Frequency Stability

There are two types of variation of pulse-repetition frequency to consider: slow drifts, which take place at a rate less than about one cycle per second, and more rapid variations. A slow drift of frequency is equivalent to detuning all circuits in the same direction by an amount equal to the frequency departure. To keep alignment displacement and the phase deviations produced by changes of pulse pattern to within tolerable bounds these frequency drifts should be kept to less than about three parts in 10^4, as calculated for detuning.

Any change of pulse-repetition frequency results in a change of the total delay through a system, even with all other parameters fixed. The change of phase for a given change of frequency is given by

$$\Delta\Phi = 2NQ\frac{\Delta f_r}{f_r} \tag{17}$$

if the timing filters are simple tuned circuits and if the rate of change is low enough to be passed by all of these filters in tandem. Since there may be a range of rates low enough to be passed by the timing filters but high enough to interfere with the signal being transmitted, it is evident that frequency stabilization is indicated. The pulse-repetition frequency is completely determined at the transmitting terminal and therefore can and should be kept from varying by more than about one part in 10^6. The short-time stability should be better; this should eliminate frequency deviations from consideration.

V. REQUIRED SYSTEM PERFORMANCE

5.1 General Requirements

We are considering in this paper the possibility of transmitting broadband signals over long distances. For purpose of illustration, let us assume that we are interested in a system which is capable of transmitting two television channels on one RF carrier. It appears likely

that such transmission would require a 20-mc sampling rate and eight digits per sample, resulting in a pulse-repetition frequency of 160 mc. For present purposes we need have only a moderately accurate estimate of such requirements, since any general conclusions we draw would be unaffected by reasonable deviations from the 160-mc figure. Such a system would be capable of transmitting other types of signal; for example, it might accept and deliver signals in the same form in which they are transmitted over the L3 coaxial cable. The L3 system, which is coming into use in the telephone plant, is described in Ref. 6.

In addition to the granularity noise resulting from quantization, the signal recovered at the end of a PCM system will contain noise resulting from errors in pulse pattern produced during transmission through the system. If the increment in time deviation at a single repeater, resulting from noise or other timing deficiencies, becomes large, there is a possibility that the resultant misalignment between signal pulses and timing pulses will result in errors in the transmitted pulse pattern. We have built an experimental system[4] employing simple timing circuits and found that the errors produced in this system by the effects of random noise on timing were entirely negligible in comparison with the number of errors produced by amplitude effects. Since this should be the case in general, we shall limit our consideration in this discussion to the effects of time deviations, or weave, upon the recovered signal.

5.2 *Over-all Effect of Time Displacements*

In operation, a PCM or other binary system transmits information sample by sample. Normally, these samples are taken at perfectly regular intervals and should be recovered in the same manner. Transmission through a system in which the timing wave suffers deviations in phase due to noise or other defects is equivalent to transmission through a system which includes a varying amount of delay. The spacing between recovered samples varies in accordance with the time deviations. If the rate of deviation is considerably less than the frequency of the recovered signal, the distortion takes the form of pure phase modulation of this signal. Fig. 9 shows the effect of time deviation upon a sine wave which has been coded, transmitted and decoded. The recovered wave is seen to be distorted. The effects of time deviations can best be determined by considering various types of signals separately. Frequency-division voice multiplex and color television appear to place the most stringent requirements on timing. They are the only types of signals which are considered in detail here. The requirements imposed by other types of signals can be calculated by similar procedures.

5.2.1 *Color Television*

Color TV has, in addition to those of black-and-white TV, a requirement that transmission through the system must not cause any appreciable change in the phase relationship between the color carrier and the reference carrier derived from the color bursts. These two carriers are at the same frequency (3.58 mc). Although the color carrier and the color bursts would be subjected to the same type of time deviation during transmission, the results would be different in the final TV receiver. To obtain reductions in effects produced by noise or interference, the reference carrier in a TV receiver is derived from a circuit having a time constant so long that, from our standpoint, we can consider the phase of its wave to be fixed. The color carrier, on the other hand, must be derived from a circuit which allows rapid variations of phase. Therefore rapid variations of delay cause corresponding changes of the phase relationship between these two 3.58-mc carriers. It is con-
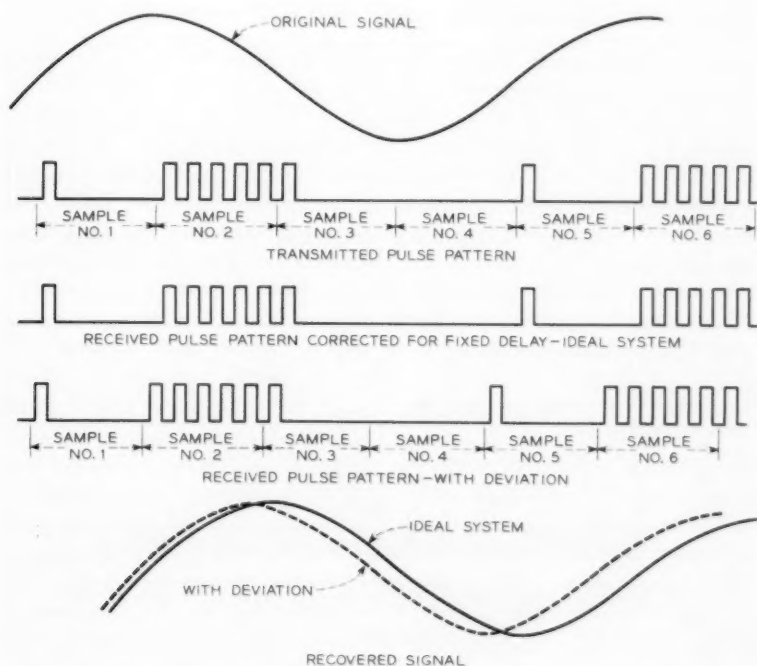

Fig. 9 — The effect of timing-wave phase deviations on a transmitted signal.

venient to express system requirements in terms of allowable phase deviations at pulse-repetition frequency. Change of delay, in milli-microseconds, is common to all frequencies being transmitted through the system, but resultant change of phase, in degrees, is proportional to frequency, so that

$$\frac{\Phi_r}{\Phi_{2.58}} = \frac{f_r}{3.58 \times 10^6},$$

where $\Phi_{3.58}$ is the change of phase of the 3.58-mc signal and Φ_r is the corresponding change of phase at pulse-repetition frequency. For a repetition frequency of 160 mc,

$$\Phi_r = 44.7 \, \Phi_{3.58}.$$

In other words, for every degree of phase deviation permitted to the color carrier 44.7 degrees, or about 0.8 radian, are permitted to the 160-mc pulse carrier.

5.2.2 L3 Frequency-Division Speech Multiplex

After decoding and filtering, the recovered signal for this system has a spectrum extending approximately from 300 kc to 8 mc. This spectrum was originally derived by combining a large number of voice channels spaced 4 kc apart and each occupying a band of about 3 kc. To become useful signals, each of these bands must be filtered out of the complete assemblage and moved back down to the audible frequency range. For the present, our interest is in the composite signal before this final filtering [see Fig. 10(a)]. If there is time displacement of samples in the course
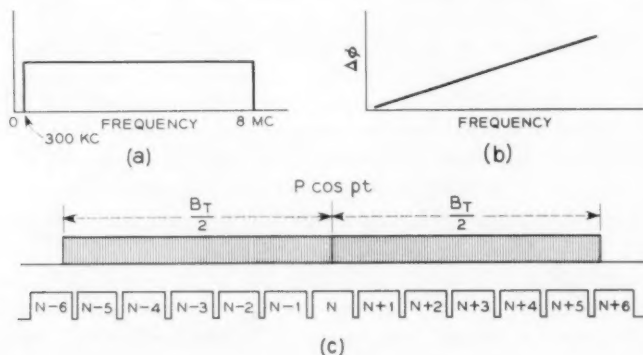

Fig. 10.— The L3 multiplex system: (a) baseband frequency range; (b) phase deviation produced by a given time deviation; (c) noise spectrum.

of transmission through the system, each component of the composite signal will be displaced in time by the same amount. Each component will thereby be subjected to a spurious phase modulation, the magnitude of which will depend linearly upon the frequency of the component; i.e., there will be a triangular distribution of phase modulation as shown by Fig. 10(b). Whether the resultant sidebands produce noise in the channel itself or crosstalk into adjacent channels depends upon their spacing from the signal. In Appendix C it is determined that the maximum rms phase deviation, Φ_r, which can be tolerated at a pulse-repetition frequency of 160 mc and still meet the assumed multiplex requirements is equal to 0.177 radian. W. R. Bennett[2] and E. D. Sunde have also analyzed this problem and arrived at substantially the same value of tolerable deviation.

We have seen that, for color TV, we can tolerate about 0.8 radian of timing-wave phase shift for every degree of allowable phase shift of the color carrier. For L3 voice multiplex we can tolerate only about 0.20 radian of timing wave phase fluctuation. Then, unless the color carrier phase must be held closer than one degree, it appears that color TV transmission requirements are much less stringent than those for voice multiplex. However, we cannot ignore the color TV requirements until we assure ourselves that this signal does not produce more phase deviation due to detuning, amplitude-to-phase conversion, etc., than is produced by the multiplex signal.

VI. ATTAINING REQUIRED SYSTEM PERFORMANCE

Since timing wave phase deviations accumulate along a chain of repeaters, it is evident that regeneration cannot be repeated indefinitely. The question then appears to be: "Is it practical to build equipment which will regenerate a few hundred times and still provide an output signal which will meet the rather stringent requirements discussed above?" Consider first the effects of random noise: To avoid an excessive number of errors due to purely amplitude effects it is necessary to maintain a signal-to-noise ratio of about 23 db* at each repeater for a chain of a few hundred repeaters. If we choose as a typical, practical system, one with a 320-mc RF bandwidth providing a 23-db S/N ratio and with timing circuits each having a Q of 200 at a pulse repetition frequency of 160 mc, the timing wave phase deviations due to random noise in 200 repeaters can be calculated from (15) to be $\Psi_{rms} = 0.056$

---

* This figure, which is the ratio of RF power at the peak of a pulse to rms noise power, is based on experimental results involving incomplete regeneration. An improved regenerator might require only a 19- to 20-db ratio.

radian, which is below the L3 voice multiplex requirements which we assumed.

We have seen that phase deviations may also arise as a result of changes of pulse pattern. Consideration of the effects of changes of pulse pattern leads us to believe that a simple timing circuit will meet system requirements if these effects exist alone. However, since detuning, amplitude-to-phase conversion and random noise will probably exist simultaneously, it may be highly desirable to obtain some improvement in performance over that provided by the simple circuits. This indicates a need for circuits which provide higher Q without sacrifice of stability. Fortunately such circuits exist and are not unduly complicated. For example, a phase-locked oscillator can be made to be the equivalent of a stable, high Q tuned circuit. Furthermore, only a limited number of these exalted-Q circuits need be employed in a system to obtain considerable advantage, as can be seen from the following example. As a reference system, assume 100 repeaters, each with a timing circuit having a Q of 100. If, at the output of the system, we insert one timing circuit with a Q of 10,000 we obtain an improvement in timing signal-to-noise ratio of about 10.5 db. If 10 of the narrow-band circuits are distributed throughout the system the improvement increases to approximately 16 db. If all repeaters are provided with the sharp circuit the total improvement is about 20 db. A practical arrangement appears to be to use a simple type of timing circuit at all unattended repeaters, with much narrower circuits at those which are attended.

VII. CONCLUSIONS

In the timing portion of a chain of regenerative repeaters (whether self-timed or externally timed) noise builds up in much the same manner as in an analog transmission system. For the purpose of analysis, this portion of a system can then be considered as a separate analog channel. The major disturbing effect of noise is the production of phase deviations of the timing wave, which in turn result in deviations of the times of occurrence of the output signal pulses.

External timing, when compared with self timing, has some definite advantages in certain systems but has disadvantages in others. Whether it is preferable to employ self timing or external timing in a particular system will depend upon many considerations beyond the scope of this paper, including the number of channels being transmitted and the delay stability of the transmission medium.

Indications are that, for long chains of repeaters, other effects, such

as those due to changes of pulse pattern, may be more detrimental to timing than the effects of random noise. Although noise and the effects of changes of pulse pattern on timing limit the number of repeaters of a given design which can be operated in tandem, it appears practicable to extend systems to a length of several hundred repeaters. Although most system requirements might be met by employing simple tuned circuits as timing filters it may prove to be preferable to obtain the considerable improvement which results from addition of a relatively few circuits of a different type which provide decreased bandwidth without sacrifice of tuning stability.

VIII. ACKNOWLEDGMENTS

W. M. Goodall first pointed out the possibility of treating the timing part of a chain of repeaters as a separate analog channel. The writer wishes to thank H. T. Friis, W. R. Bennett and H. E. Rowe for the use as background of material which is not yet published as well as for the specific material mentioned here. He is also indebted to J. C. Schelleng and others for many helpful suggestions in the preparation of this paper.

APPENDIX A

Random Noise in a Single Repeater with a Square-Law Detector

As a starting point, we assume that all pulses are present and that the RF filtering is so adjusted that the RF signal into the detector can be expressed as

$$f(t) = \frac{A}{2} (1 + \cos 2\pi f_r t) \cos 2\pi f_c t + V_N. \tag{18}$$

The signal function is shown on Fig. 6(b). The corresponding envelope is indicated by Fig. 6(a) and the frequency spectrum by Fig. 6(c); V_N represents the voltage arising from noise which is assumed to have a uniform power density of W_0 watts per cycle of bandwidth. If we follow the procedure of S. O. Rice[7] and assume the detector to have a current-voltage characteristic given by

$$I = \alpha V^2, \tag{19}$$

we obtain the detector output currents by squaring (18) and multiplying by α.

The recovered timing wave (at frequency f_r) is given by

$$V_t = \frac{\alpha A^2}{4} \cos 2\pi f_r t, \tag{20}$$

and, for the timing wave power,

$$W_s = \frac{\alpha^2 A^4}{32}. \tag{21}$$

An expression for the noise spectrum at the output of the detector is given by Rice's equation (4.5–17) in Ref. 7. If we assume that our signal-to-noise ratio is large, we can neglect the cross-products among noise components in comparison to the cross-products between signal components and noise components. Since our timing circuit selects only those noise components which lie in a narrow band about f_r at the output of the detector, we need consider only the corresponding terms in Rice's equation. We find that the timing noise power at the output of the filter is

$$W_{na} = \frac{\alpha^2 A^2 W_0 B_t}{2}, \tag{22}$$

where B_t is the effective bandwidth of the timing filter. The timing wave signal-to-noise ratio is

$$\frac{W_s}{W_{na}} = \frac{A^2}{16 W_0 B_t}.$$

Since the RF signal power, P_s, at the peak of a pulse is equal to $A^2/2$, this equation can be written

$$\frac{W_s}{W_{na}} = \frac{P_s}{8 W_0 B_t}. \tag{23}$$

APPENDIX B

Random Noise — Chain of Repeaters with Tuned Circuits

From the standpoint of timing a chain of self-timed repeaters is equivalent to the chain of tuned circuits of Fig. 11. There will be unity gain, at midband, from one circuit to the next and noise arising at the input to each of the amplifiers shown.

H. T. Friis has shown that, for a chain of m such tuned circuits in tandem, the effective bandwidth is given by

$$B_{\text{eff}} = \frac{\pi f_r}{2Q} h_m, \tag{24}$$

where

$$h_m = \frac{(2m)!}{(2^m \times m!)^2} \quad \text{and} \quad Q \gg 1. \tag{25}$$

The first five values of h_m calculated from (25) are:

m	1	2	3	4	5
h_m	1	0.5	0.375	0.312	0.27

For large values of m,

$$h_m \doteq \frac{1}{\sqrt{\pi m}}. \tag{26}$$

Noise arising in the last amplifier, or repeater, of the chain will be filtered by one tuned circuit, that arising at the next-to-the-last amplifier will be filtered by two circuits and so on to noise at the input to the chain which will be filtered by all N circuits. The total noise power contributed by the N repeaters is given by

$$(W_{na})_N = \frac{\pi f_r}{2Q} \left(\sum_{m=1}^{m=N} h_m \right) d_0, \tag{27}$$

where d_0 is the noise power density at the input to a single amplifier.

If we use the simpler expression for h_m as given in (26),

$$(W_{na})_N \doteq \frac{\pi f_r}{2Q} \left(\sum_{m=1}^{m=N} \frac{1}{\sqrt{\pi m}} \right) d_0. \tag{28}$$

Since (26) gives a good approximation to h_m only for values of m greater than about five, we can increase the accuracy of (28) by subtracting out the first five terms and replacing them by the values of h_m obtained
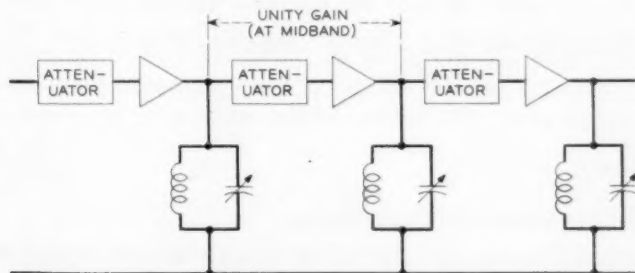

Fig. 11 — Simple tuned circuits in tandem.

from the above table. Equation (28) then becomes

$$(W_{na})_N = \frac{\pi f_r}{2Q}\left[\frac{1}{\sqrt{\pi}}\left(\sum_{m=1}^{m=N}\frac{1}{\sqrt{m}} - \sum_{m=1}^{m=5}\frac{1}{\sqrt{m}}\right) + 2.46\right]d_0$$
$$= \frac{\pi f_r}{2Q}\left[\frac{1}{\sqrt{\pi}}\left(\sum_{m=1}^{m=N}\frac{1}{\sqrt{m}} - 3.23\right) + 2.46\right]d_0. \quad (29)$$

Since

$$\sum_{m=1}^{m=N}\frac{1}{\sqrt{m}} = 2\sqrt{N} - 1$$

to a fair degree of accuracy, (29) becomes

$$(W_{na})_N \doteq \frac{\sqrt{\pi}\, f_r\, \sqrt{N}}{Q}\, d_0 = \frac{1.77 f_r\, \sqrt{N}}{Q}\, d_0. \quad (30)$$

From (22) of Appendix A we see that the appropriate value of d_0 is

$$d_0 = \frac{\alpha^2 A^2 W_0}{2}.$$

Then,

$$(W_{na})_N = \frac{1.77\alpha^2 A^2 W_0 f_r \sqrt{N}}{2Q}. \quad (31)$$

The timing-wave power is the same at the end of the chain as for a single repeater. From (21) of Appendix A,

$$W_s = \frac{\alpha^2 A^4}{32},$$

and

$$\left(\frac{W_s}{W_{na}}\right)_N = \frac{A^2 Q}{2 \times 14.2 W_0 f_r \sqrt{N}} = \frac{P_s Q}{14.2 W_0 f_r \sqrt{N}}. \quad (32)$$

If we consider only the quadrature component of the noise,

$$\left(\frac{W_s}{W_n}\right)_N = \frac{P_s Q}{7.1 W_0 f_r \sqrt{N}}. \quad (33)$$

APPENDIX C

Required Performance for Voice Multiplex

In this section, the maximum tolerable phase deviation of the timing wave at the end of a chain of repeaters is calculated from parameters based on the L3 system.

As an aid to understanding the effects of time deviations let us first consider the case where a single audio tone is being transmitted through

one voice channel. After decoding and filtering, this tone will result in a single frequency lying somewhere in the range between 300 kc and 8 mc (see Fig. 10). If the samples from which this frequency is derived vary in time the component will be phase-modulated. The spacing of the resultant sidebands from the channel frequency is determined by the rate at which its phase is being deviated. Since we are interested in determining results for the voice channel which is most affected, we choose a channel at the upper end of the band but far enough inside the band that noise sidebands produced when it is active do not exceed 8 mc. Let the decoded signal component resulting from the tone be represented by $P \cos pt$. If the phase deviation is sinusoidal, as it will be if produced by a single interfering wave somewhere in the system, the decoded output will be:

$$B(t) = P \cos (pt + \Phi \cos qt), \qquad (34)$$

where Φ is the magnitude of the phase deviation and q is the angular frequency at which this deviation takes place. By expanding (34) and limiting consideration to the case where the deviation is small, the approximate value of signal is:

$$V(t) \doteq P \cos pt - \frac{P\Phi}{2} [\sin (p + q)t + \sin (p - q)t]. \qquad (35)$$

Thus, $V(t)$ is seen to consist of the signal, $P \cos pt$, plus two sidebands spaced from it by a frequency corresponding to q. The signal power in a single channel carrying the tone $P \cos pt$ is $P^2/2$. The power in the two sidebands expressed by (35) is

$$\left(\frac{P\Phi}{2}\right)^2 = \frac{P^2\Phi^2}{4}. \qquad (36)$$

When the phase deviations are produced by noise or other random effects each signal will have sidebands distributed about it and extending over the band B_T, as shown by Fig. 10(c), where B_T is the effective bandwidth of all of the timing circuits in tandem. Fig. 10(c) shows the sideband power to be evenly distributed over H voice channels, where H is the number of voice bands encompassed by the timing-circuit band. Thus,

$$H = \frac{B_T}{B_s}, \qquad (37)$$

where B_s is the bandwidth of a signal channel.

If a number of adjacent channels are active simultaneously, each will have a spectrum as shown by Fig. 10(c). These spectra will overlap in frequency, as shown in Fig. 12. For the sake of simplicity, we first assume that all channels are equally loaded. It is evident from Fig. 12

that a total of H channels will throw interfering power into each voice channel. With the interference from each channel spread over H channels and with H channels contributing to each other, the end result is the same as though all of the interfering power produced by a channel were concentrated in a single channel. If T is the maximum tone carrying capacity of a voice channel, the maximum signal-to-noise ratio is obtained by dividing the maximum signal power, $T^2/2$, by the noise sideband power $(P^2\Phi^2)/4$, as expressed by (36). We obtain

$$\frac{S}{N} \text{ (audio)} = \frac{2T^2}{P^2\Phi^2}. \tag{38}$$

Let us assume that multiplex requirements call for a signal-to-noise ratio of at least 60 db or 10^6. Then:

$$\Phi^2 = \frac{2T^2}{10^6 P^2}, \qquad \text{or} \qquad \Phi = \frac{\sqrt{2} \times 10^{-3} T}{P}. \tag{39}$$

If all channels are loaded to capacity with tone, $P = T$ and $\Phi = \sqrt{2} \times 10^{-3}$. This is the peak value of the maximum tolerable sinusoidal deviation. The corresponding rms deviation, whether sinusoidal or random, is 1×10^{-3} radian.

When the channels are loaded with speech, as they normally will be, the interference will be less than when they are loaded with tone. For speech loading the interference produced by each channel will still be
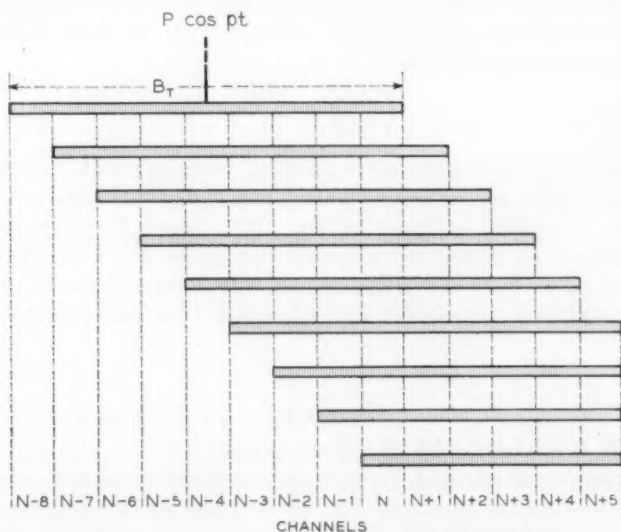

Fig. 12 — Multiplex system—overlapping noise spectra.

distributed over H channels, but the total interfering power will no longer be H times that produced in a single channel. This is the well-known speech multiplexing effect. At any instant the different channels will produce different amounts of disturbance, so that the total peak noise sideband power falling into a single channel is given by

$$\text{Peak noise power} = \frac{1}{H} \sum_{m=1}^{m=H} \frac{P_m^2 \Phi^2}{2} = \frac{\Phi^2}{2H} \sum_{m=1}^{m=H} P_m^2. \qquad (40)$$

To calculate H, we first determine the effective bandwidth of all of the timing channels in tandem. It is evident from (30) of Appendix B that, for the case of tuned circuits,

$$B_T = \frac{1.77 f_r}{Q \sqrt{N}}. \qquad (41)$$

As an example, consider a chain of 100 repeaters, each with a timing circuit Q of 100 and operating at a 160 mc repetition rate. For this system, $B_T = 284$ kc. For a voice-channel spacing of 4 kc, $H = 284/4 = 71$ channels. From the Holbrook-Dixon multiplex data[8] we find that 71 voice channels require approximately 10 db more peak power-handling capacity than is required by a single channel. Then,

$$\sum_{m=1}^{m=71} P_m^2 = 10 P^2,$$

$$\text{Total peak power noise} = \frac{\Phi^2 \times 10 P^2}{2 \times 71}.$$

Using this expression for the noise power,

$$\frac{S}{N} \text{ (peak audio)} = \frac{14.2 T^2}{\Phi^2 P^2}. \qquad (42)$$

To obtain the rms signal-to-noise ratio, we must take into account the tone peak factor of 3 db and the speech peak factor of 13.6 db for 71 channels, as obtained from the Holbrook-Dixon data. Then,

$$\frac{S}{N} \text{ (rms audio)} = \frac{156 T^2}{\Phi^2 P^2}. \qquad (43)$$

As before in determining the required signal-to-noise ratio, we take $T = P$ and assume that the required power ratio is 60 db or 10^6. Setting the ratio as expressed by (43) equal to 10^6, we find $\Phi = 12.5 \times 10^{-3}$ radian at 8 mc, and

$$\Phi_{rms} = 8.84 \times 10^{-3} \text{ radian at 8 mc}$$
$$= 20 \times 8.84 \times 10^{-3} = 0.177 \text{ radian at 160 mc.} \qquad (44)$$

SYMBOLS

s = peak amplitude of sinusoidal timing wave.

W_s = mean power of the recovered sinusoidal timing wave.

δ = peak amplitude of a single interfering component whose frequency is near that of the timing wave.

δ_l = peak amplitude of each of the phase modulation sidebands resulting from limiting the timing wave and a single interference of peak amplitude δ.

W_{na} = mean noise power at the output of the timing filter and before limiting.

W_n = mean noise power at the output of the timing filter after limiting.

P_s = mean RF power at the peak of a signal pulse.

A = peak RF amplitude corresponding to P_s.

W_0 = RF noise power density in watts per cycle of bandwidth.

B_t = effective bandwidth of the timing circuit of a single repeater.

B_T = effective bandwidth of all of the timing circuits in a system operating in tandem.

B = bandwidth of RF section of repeater employing ideal flat filters.

M = the ratio of the total number of pulse positions contained in a given time interval to the number of pulses present in the same interval.

f_c = frequency of an RF carrier.

f_r = pulse-repetition frequency (usually considered to be 160 mc in this article).

N = number of repeaters in the chain of interest.

Ψ_{max} = maximum instantaneous phase excursion of the sinusoidal timing wave from its long time average value.

ϵ = rms phase deviations which result when a random pulse pattern excites a detuned circuit.

β = steady-state phase shift resulting from detuning of a tuned circuit.

k = ratio of the amount of detuning in cycles per second to the nominal resonant frequency of the circuit.

REFERENCES

1. Sunde, E. D., Self-Timing Regenerative Repeaters, B.S.T.J., **36**, July 1957, pp. 891–938.
2. Bennett, W. R., this issue, pp. 1501–1542.
3. Rowe, H. E., this issue, pp. 1543–1598.
4. DeLange, O. E. and Pustelnyk, M., this issue, pp. 1487–1500.
5. Oliver, B. M., Pierce, J. R., and Shannon, C. E., The Philosophy of PCM, Proc. I.R.E., **36**, November 1948, pp. 1324–1331.
6. Elmendorf, C. H., Ehrbar, R. D., Klie, R. H. and Grossman, A. J., B.S.T.J., **32**, July 1953, pp. 779–1005.
7. Rice, S. O., Mathematical Analysis of Random Noise, B.S.T.J., **24**, January 1945, pp. 46–153.
8. Holbrook, B. D. and Dixon, J. T., Load Rating Theory for Multi-Channel Amplifiers, B.S.T.J., **18**, October 1939, pp. 624–644.

Experiments on the Timing of Regenerative Repeaters

By O. E. DE LANGE and M. PUSTELNYK

This paper describes a number of experiments performed with self-timed binary regenerative repeaters to determine the behavior of the timing portion of a chain of such repeaters.

I. INTRODUCTION

Three other papers[1,2,3] appearing in this issue of the Bell System Technical Journal discuss and analyze the timing of regenerative repeaters. It is the purpose of this paper to describe some experiments performed on repeaters of the same general type to determine their performance in the presence of random noise.

Although our interest is largely in microwave systems, the repeaters for such systems are relatively complicated and expensive. Thus, it is impractical to build a long chain of them expressly for experimental purposes. The recovered timing information is at baseband frequency in a carrier system as well as in a baseband system. For these reasons, we built a chain of baseband repeaters to simulate, as nearly as possible, the microwave repeaters of interest. A pulse-repetition frequency of 10 mc was chosen. The experiments had two objectives: first, to determine the effects of noise in producing pulse errors through its action on the timing circuit and second, to determine the effects of noise in producing time or phase deviations of signal pulses, as discussed in Refs. 1, 2 and 3. From these results, we hope to determine what characteristics the timing circuits must have in order to meet the over-all system requirements. The experiments concern only self-timed repeaters.

It was found that, with a fixed pulse pattern and the simple timing circuits employed, the number of errors produced by the effects of noise upon timing was negligibly small in comparison to the number produced by other effects of noise. This was true even when the system was operating under very adverse conditions. It might be expected from

these results that random noise in the timing circuit will not be a major cause of errors in an actual system. The results also indicate that, as far as random noise is concerned, the amount of time deviation of pulses at the output of a long chain of repeaters can be kept within tolerable bounds.

II. SCOPE OF THE EXPERIMENTS

Self-timed regenerative repeaters in general appear to have a common defect in that they convert changes of pulse pattern into changes in the time of occurrence of the pulses out of the regenerator. When a PCM system is transmitting information the pulse pattern changes constantly; therefore, if the system is to be useful its timing circuits must function satisfactorily in the presence of such changes. The degree of conversion depends upon the characteristics of the particular repeater employed and probably can be reduced to a tolerable value by proper design. However, before one decides to instrument a timing circuit which will tolerate a changing pulse pattern it is desirable to obtain an answer to the following question: Aside from the effects of pulse pattern changes, can a system be built which will function in the presence of random noise, and, if so, what are the requirements placed on the equipment by noise considerations alone? After this study has been made we can then face the practical problem of designing a timing circuit which will be sufficiently free of the defects of instrumentation to permit satisfactory operation of the over-all system. In order to avoid the effects of dynamic changes of pulse pattern, the experiments described here were performed with repetitive patterns.

III. TIMING ERRORS

By "timing errors" we mean changes in pulse pattern produced during transmission by imperfections in the process of pulse timing. For example, if one pulse is lost from a time slot which should contain a pulse this constitutes one error. In our experiments the number of errors was determined by comparing the pulse pattern at the input to the system with the pattern at the output of the system, pulse by pulse and space by space. Whenever there was a difference between the two patterns the comparator, which consisted of an "exclusive OR" circuit, emitted a pulse. These pulses were counted by a high-speed binary counter.

Timing errors result from misalignment between signal pulses and timing pulses. It is obvious that, if a timing pulse is displaced by a

sufficient amount, it will occur at a time when the corresponding signal pulse is below standard amplitude and there will be no output from the regenerator. Even when the displacements alone are not great enough to produce errors they can increase the probability of errors due to amplitude effects. The number of errors produced in any given case will depend upon the widths and shapes of both signal and timing pulses as well as upon the amount of misalignment. Because of the number of factors influencing the production of timing errors, it is felt that experimental results obtained from a typical repeater might be more meaningful than those obtained from an analysis based on uncertain assumptions.

3.1 *Comparison of Timing and Amplitude Errors — Single Repeater*

The first experiment involved only a single repeater, a block diagram of which is shown in Fig. 1. The signal path is seen to consist of a filter, a baseband amplifier, the regenerator, another filter and an amplifier stage to couple the output to the succeeding repeater when the unit is part of a chain. The timing path consists of an amplifier, a tuned circuit, a limiter, a second amplifier and a pulse generator. Means are provided for adding broadband random noise from an external source to each repeater. The repeaters are set up in such a way that this noise can be inserted in the timing path only, in the slicer circuit only or in both simultaneously.
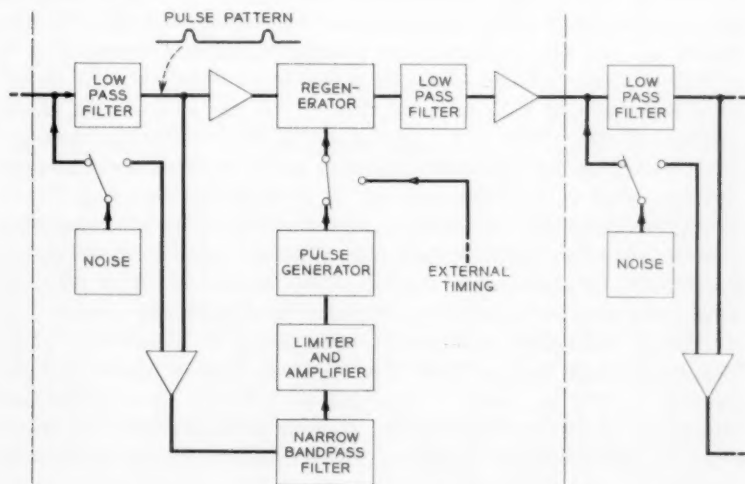

Fig. 1 — Block diagram of the baseband repeater.

The regenerator is essentially an AND gate; i.e., there is an output pulse only when there is a pulse on both of two inputs simultaneously. One input consists of a train of timing, or "clock" pulses. The other input is the train of signal pulses which are to be regenerated. The timing pulses are about one third as long as the signal pulses, so that the time of occurrence of each output pulse is determined almost entirely by the time of occurrence of a corresponding timing pulse. Ideally, there would be an additional requirement placed on the regenerator: that it respond completely to voltages in the signal path with amplitudes greater than one half of normal peak pulse amplitude and respond not at all to voltages less than this value; i.e., it would provide an ideal slicer characteristic. However, the microwave repeater which is being simulated (described in Ref. 4) has a slicer characteristic which departs considerably from the ideal. The baseband slicer characteristic was made to match that of the microwave repeater. The regeneration is performed by a 6AS6 pentode vacuum tube with the signal pulses applied to its suppressor and the timing pulses to its control grid. Since the type of baseband repeater employed was designed solely for these experiments and would never be used in an actual system it will not be described in greater detail.

For most of the experiments, the input pulse pattern was adjusted to have only one pulse present for every nine possible pulse positions in order to set up very unfavorable conditions for timing. With noise applied to the main input of the repeater so as to affect both the timing circuit and the slicing circuit, the number of errors corresponding to each of a number of signal-to-noise ratios was determined. The experiment was repeated with noise applied only to the timing path of the repeater. Finally, with noise applied to the main input, the repeater was timed from an external, noise-free source so that noise affected only the slicer path of the repeater. It is apparent from Fig. 2 that, for a given signal-to-noise ratio,* many more errors are produced when noise is applied to the slicer only than when it is applied to the timing circuit only. It can also be seen that the number of errors produced when noise acts on both circuits together is considerably greater than the sum of the errors produced by noise acting on each circuit alone. This results from the fact that less amplitude noise is required to produce an error when there is misalignment between signal pulses and timing pulses. Although the number of errors produced by timing effects alone is very small in comparison to the number produced by noise

---

* We define signal-to-noise ratio as the ratio of peak pulse power to mean noise power.
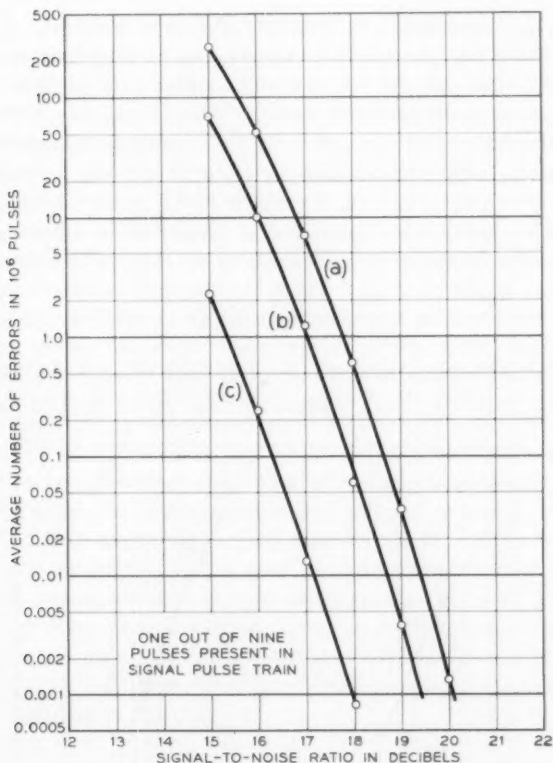
Fig. 2 — Average pulse error rate with linear timing amplifier: (a) self-timed, noise into both slicing and timing paths; (b) externally timed, noise into slicing path only; (c) self-timed, noise into timing path only.

acting on the slicer alone, timing noise is still exerting considerable influence on the total number of errors. It should be pointed out again that the conditions under which these counts were made are very adverse to timing; for example, an increase in the number of signal pulses to two out of a possible nine would be equivalent to improving the signal-to-noise ratio by 6 db, which would make timing errors entirely negligible.

For the above experiments, the timing circuit suffered a severe penalty from the fact that noise was applied to the timing filter continuously, whereas signal energy was applied to this circuit by only one out of each nine possible pulse positions. As a result, the component of

the signal at pulse-repetition frequency had only one-ninth of the amplitude it would have had with all pulses present. In a later experiment, this penalty was reduced by inserting a nonlinear element between the main pulse path and the tuned circuit. This was accomplished by back-biasing one of the tubes in the timing amplifier so that it conducts only when the voltage on its grid exceeds some threshold value. In this way, most of the noise which occurs between signal pulses is eliminated, thus removing a considerable amount of noise power from the timing circuit when there are few signal pulses present. Comparing Fig. 3 with Fig. 2 makes it evident that the nonlinear amplifier has produced an apparent increase of about 3 db in the timing signal-to-noise ratio. Other experiments with the same equipment produced improvements of approximately 5 db. One might expect that such a peak amplifier would reduce the noise power applied to the tim-
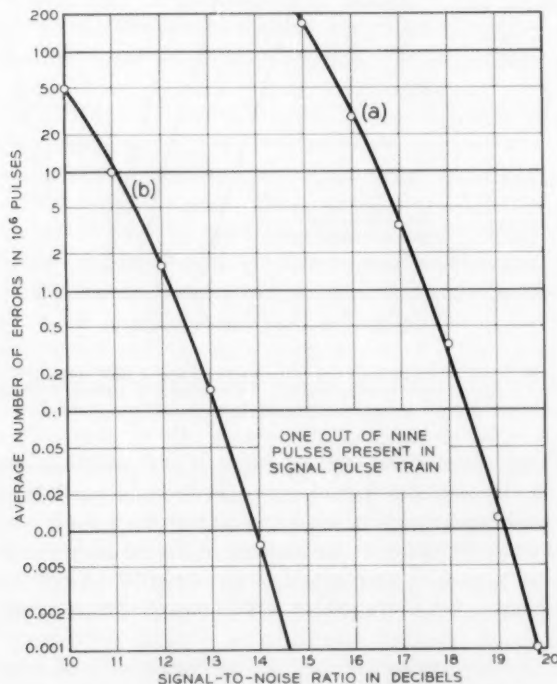

Fig. 3 — Average pulse error rate for a self-timed repeater with a nonlinear amplifier: (a) noise into both slicing and timing paths; (b) noise into timing path only.

ing circuit by a factor of nine to produce an improvement of 9 db. However, the reduction of noise power is accompanied by a redistribution of the noise frequency spectrum, which probably accounts for the fact that the improvement is less than might otherwise be expected. With the improvement obtained by use of the nonlinear circuit, it appears safe to assume that, for a single repeater employing a suitable timing circuit of simple design, the number of errors produced by the effects of random noise upon timing is completely negligible at all times in comparison with the number produced by other effects of noise.

3.2 *Comparison of Errors — Chain of Repeaters*

This experiment was performed to determine the manner in which the number of errors increases as the number of repeaters is increased. For this case, the signal-to-noise ratio at the input to each of five repeaters was held constant at 22 db. The pulse errors produced by employing first one, then two and so on up to five repeaters were counted. The timing filters for these experiments consisted of tuned circuits with an effective Q of about 80 in each repeater. It is obvious from Fig. 4 that, for five repeaters, the number of errors produced by the effects of noise on timing alone is very small in comparison with the number produced by other effects of noise. These results were obtained with the linear timing amplifier. With the nonlinear circuit, the number of errors produced by timing alone would have been much too small to measure with the equipment available.

Some discussion of the results indicated by Fig. 4 is needed at this point. Although these experiments were primarily directed toward the study of timing, we obtained as by-products some interesting results concerning the amplitude effects of noise. For ideal regeneration of pulses, the total number of errors produced in a system should vary almost linearly with the number of repeaters involved since, for such a system, no noise would be passed from one repeater to the next. For the system under discussion the regenerators do not have ideal input-amplitude vs. output-amplitude characteristics, and therefore some noise effects are passed from one repeater to the next.* If only the first repeater of the chain had noise at its input the noise power (except for peaks exceeding the slicing level) would be reduced at each successive repeater to such an extent that it would be negligible after traversing four or five of the partially regenerating repeaters. However, when each repeater has noise at its input there is a rapid build up of error rate for

---

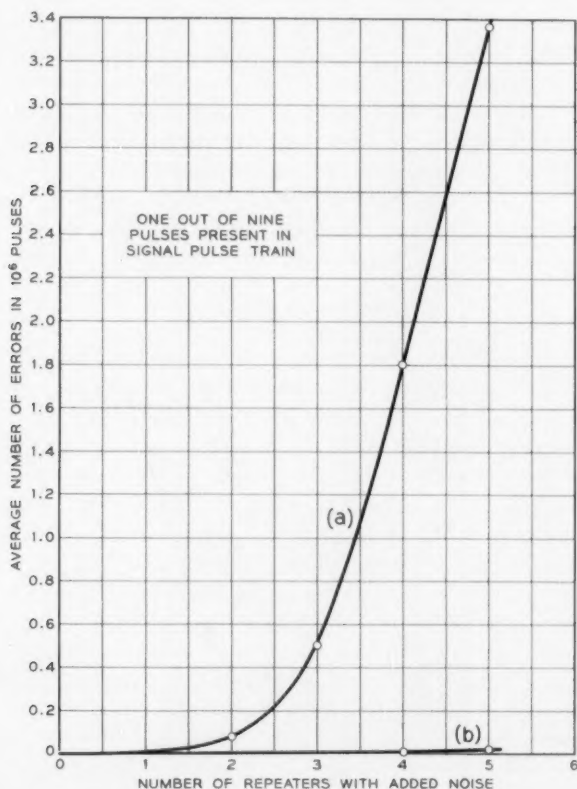* For a discussion of this "partial regeneration" of pulses see page 17 of Ref. 4.

Fig. 4 — Error build up in a chain of self-timed repeaters: (a) noise into both timing and slicing paths; (b) noise into timing path only.

the first four or five repeaters, after which the rate of increase stabilizes, as shown on Fig. 4. From this point on the number of errors increases nearly linearly with the number of repeaters. The net result of partial amplitude regeneration appears to be a reduction in the effective signal-to-noise ratio in the main signal path. This should have little effect on timing. The noise which is passed from one repeater to the next is re-timed in the regenerator in much the same way as signal pulses. We have found in our experimental systems that, for a long chain of repeaters, the signal-to-noise ratio required for a given error rate is three to four db greater than would be required for a system employing ideal regen-erators.

IV. PHASE OR TIME DEVIATIONS

Having shown that noise has negligible effect in producing pulse errors we next proceed to investigate the effect of noise in producing phase deviations of the recovered timing wave. In the accompanying papers[1, 2, 3] it is shown that, for a chain of repeaters employing tuned circuits as timing filters, the effective bandwidth decreases as the number of repeaters is increased. Then, if there is noise only at the input to the first repeater its effect should decrease at successive repeaters along the chain. That this is actually the case is shown by the following experiment: A sufficient amount of noise was added to the signal at the input to the first repeater to provide a 20-db signal-to-noise ratio. The timing wave phase deviation at this repeater was measured to be 20°. Similarly, the phase deviations at the succeeding repeaters, which were free of added noise, were measured. With the measured deviation of 20° for the first repeater as a starting point, the expected deviations at succeeding repeaters were calculated on the basis that noise power should be directly proportional to effective bandwidth. The measured and calculated deviations, which are given in Table I, are seen to agree fairly well.

One of the early experiments was performed to determine, in a qualitative way, whether or not a long chain of repeaters can be made to perform satisfactorily in the presence of noise, especially with respect to timing. A chain of 22 repeaters of the type shown on Fig. 1 was set up for this purpose. Broadband noise was added to each repeater in such a way as to affect both the slicing and timing circuits and was adjusted in level to produce a 20-db signal-to-noise ratio at the input to each regenerator.

As far as could be determined by observing its output with an oscilloscope, the system performance was very satisfactory. Various repetitive pulse patterns were set up manually and found to be transmitted without discernible errors. (Our more sensitive error-detecting equipment would probably have detected some errors.) Timing-phase devia-

TABLE I

Number of Repeaters	Calculated Deviation	Measured Deviation
1	20.0°	20.0°
2	12.3°	12.8°
3	10.5°	10.0°
4	9.5°	7.6°
5	8.8°	7.1°

tions were determined by observing the timing wave recovered by the last repeater in the chain. Deviations of this wave were practically identical with deviations of the output pulses. Fig. 5 illustrates the degree of deviation produced by this amount of noise. The upper trace shows a jitter-free reference sine wave which can be compared with the timing wave recovered from the 22nd repeater. This timing wave is shown on the bottom trace. The apparent broadening of the lower trace makes it evident that there was some phase and amplitude modulation present on the recovered timing wave. Unfortunately the phase-measuring equipment was not available at the time of this experiment; however, it is apparent from the bottom trace that the phase deviations were small. By careful measurement of the horizontal displacement of the trace, it is estimated that the rms deviation was about 4.5°. If we adopt this figure of 4.5° for 22 repeaters we can calculate the deviation to be expected for a chain of 200 repeaters on the basis that deviation varies as the fourth root of the number of repeaters (see Refs. 2 and 3). Such calculations indicate a deviation of about 8° for the long chain operating under the conditions of this experiment. These results are consistent with those calculated in Ref. 2 in indicating that, from the standpoint of noise, even simple timing circuits can meet the very stringent system requirements.

V. CALCULATION OF PHASE DEVIATIONS

To make possible comparisons between measured deviations and expected deviations of phase, the expected values are calculated here for conditions corresponding to those under which the measurements were made. Measured and calculated deviations are found to agree within the range of experimental error, both for a single repeater operat-
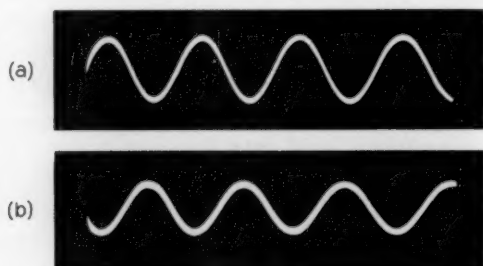

Fig. 5 — Comparison of (a) a reference 10-mc sine wave and (b) the recovered 10-mc timing wave at the output of 22 self-timed repeaters.
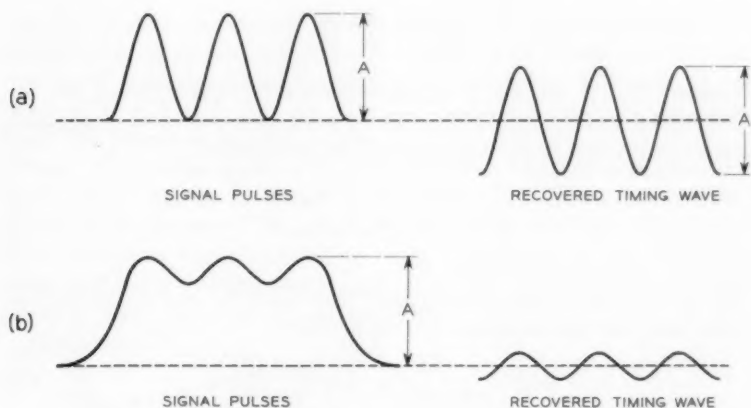
Fig. 6 — Signal pulses and recovered timing wave: (a) broadband input filter, (b) narrowband input filter.

ing under very adverse conditions and for a chain of 22 repeaters operating under more favorable conditions.

Although the method of calculation employed is basically the same as that employed in Ref. 2, it is somewhat different, because here we are dealing with systems which are completely baseband whereas the previous article considers only carrier systems. For this reason, calculations are given in some detail.

From the standpoint of filtering we choose the same conditions that were assumed in the previous article; i.e., the filtering following each regenerator is such that the signal pulses applied to the narrow filter are of raised cosine form, as shown in Fig. 6(a). If the filter at the input to each repeater has an amplitude characteristic as shown by the dashed line of Fig. 7 and a linear phase characteristic the signal pulses will not be altered in form by passage through it.* The relationship between signal pulses and recovered timing wave will be as shown on Fig. 6(a). If the noise power density is W_0 the total noise power, W_N, contained in the band B ahead of the timing filter is given by $W_N = W_0 B$. In the same way, the total noise power at the output of the timing filter is given by $W_{na} = W_0 b$, where b is the bandwidth of the timing filter. Peak pulse power W_p is A^2. The mean timing wave power is given by $W_s = A^2/8$ [see Fig. 6(a)]. From these relationships, we find

---

* This is strictly true only when all pulses are present. In any case, the frequency component at the pulse-repetition frequency of 10 mc will not be affected.

$$\frac{W_s}{W_{na}} = \frac{W_p}{W_N}\frac{B}{8b}, \tag{1}$$

where W_s/W_{na} is the timing signal-to-noise ratio after filtering but before limiting and W_p/W_N is the ratio of peak pulse power to mean noise power at the input to the timing filter. After limiting,

$$\frac{W_s}{W_n} = \frac{W_p}{W_N}\frac{B}{4b} = \frac{W_p}{W_N}\frac{Q}{2\pi}, \tag{2}$$

since $B = f_r$ and $b = \pi f_r/2Q$.

By the same procedure employed in Section 3.5 of the previous article, we find the rms phase deviation to be given by

$$\Psi_{rms} = \sqrt{\frac{W_n}{W_s}} = 2.51\sqrt{\frac{W_N}{QW_p}}. \tag{3}$$

The filter used at the input to each repeater actually has a frequency characteristic as shown by the solid line of Fig. 7, which results in signal pulses as shown by Fig. 6(b). It is evident from Fig. 7 that the additional band limiting does not affect the timing signal-to-noise ratio to any extent, since both timing signal and timing noise are recovered from the narrow band b and are reduced by very nearly the same amount by the frequency discrimination. Equation (3), with W_s/W_n defined by (2) is still accurate for the limited band. Note that W_p/W_N is the signal-to-noise ratio which would exist for the filter whose characteristic is depicted by the dashed curve of Fig. 7.
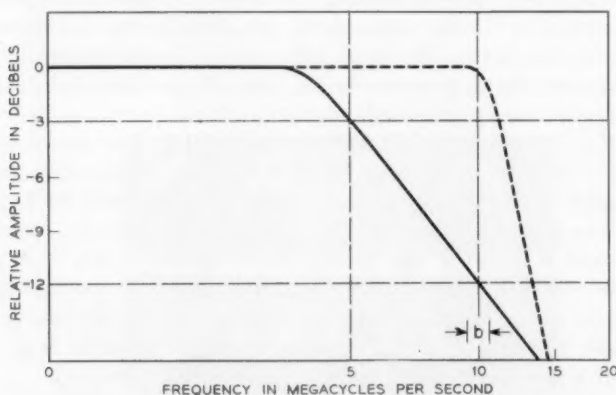

Fig. 7 — Frequency response of the lowpass input filter. The characteristic of the experimental filter is shown by the solid line. The timing filter bandwidth is shown as b.

In one experiment we operated a single repeater with a pulse signal-to-noise ratio of 20 db and with only one out of a possible nine pulses present. Input noise and signal were limited in bandwidth as shown by the solid curve of Fig. 7; i.e., the effective bandwidth was approximately 5 mc. The timing circuit had an effective Q of 80 at 10 mc.

Since (3) is based on the signal-to-noise ratio which would exist in a 10-mc band, whereas the ratio was measured in a 5-mc band, it is necessary to make a conversion to obtain the proper value to use in the equation. This value was calculated to be approximately 21 db. The ratio varies slightly with pulse pattern due to slight changes in pulse shape; however, these variations should be small.

Then, from (3),

$$\Psi_{rms} = 2.51 \sqrt{\frac{1}{80} \times \frac{1}{125}} = 2.51 \times 0.01 = 0.025 \text{ radian}$$

for the case where all pulses are present. With a linear timing input amplifier and only one out of nine pulses present, the timing wave amplitude is reduced by a factor of nine, whereas noise is unaffected by the number of pulses. The phase deviation then becomes $9 \times 0.025 = 0.225$ radian or $13°$. This is in fair agreement with the value of $17°$ measured under the same conditions.

A chain of 22 repeaters was operated under the conditions described above but with seven out of nine pulses present. Then, for the first repeater in the chain, the deviation should be $\Psi_{rms} = 9/7 \times 0.025 = 0.032$ radian. From (30) of the previous article, it is evident that the effective bandwidth of a chain of N repeaters is

$$B_{eff} = \frac{1.77 f_r}{Q\sqrt{N}}.$$

The bandwidth of a single repeater is

$$\frac{\pi f_r}{2Q} = \frac{1.57 f_r}{Q}.$$

The ratio of noise power at the output of the Nth repeater to that at the output of the first repeater is

$$\frac{(W_n)_N}{(W_n)_1} = 1.13 \sqrt{N}.$$

Then

$$\Psi_N = 1.06 N^{1/4} \Psi_1.$$

For our example:

$$\Psi_{22} = 1.06 \times 2.17 \times 0.032 = 0.074 \text{ rad.} = 4.2°.$$

This is consistent with the estimate made by means of the photograph of Fig. 6.

VI. PHASE CONVERSION

Changes of pulse pattern produced by manual switching were found to result in changes of delay, or phase, through a repeater of the type employed in these experiments. This change amounted to about 10° at the 10-mc pulse-repetition frequency when the pattern was changed from one pulse out of nine to seven out of nine. Any practical system will have to be designed in such a way that the total phase deviation resulting from expected changes of pulse pattern will not exceed system requirements. In an experimental timing circuit designed for the microwave system, phase conversion has been reduced to about one-tenth of that produced in the baseband repeaters.

VII. CONCLUSIONS

It was found that, with a fixed pulse pattern and the simple timing circuits employed, the number of errors produced by the effects of noise upon timing was negligibly small in comparison to the number produced by other effects of noise. This was true even when the system was operating under very adverse conditions. It might be expected from these results that random noise in the timing circuit will not be a major cause of errors in an actual system. The results also indicate that, as far as random noise is concerned, the amount of time deviation of pulses at the output of a long chain of repeaters can be kept within tolerable bounds.

The above experiments were performed with repetitive pulse patterns. We have not made an adequate study of the effects of rapid changes of pulse pattern upon timing. Indications are that these effects may be much more serious than the effects of random noise on timing and should, therefore, be studied more extensively.

REFERENCES

1. Bennett, W. R., this issue, pp. 1501–1542.
2. DeLange, O. E., this issue, pp. 1455–1486.
3. Rowe, H. E., this issue, pp. 1543–1598.
4. DeLange, O. E., Experiments on the Regeneration of Binary Microwave Pulses, B.S.T.J., **35**, January 1956, pp. 67–90.

Statistics of Regenerative
Digital Transmission

By W. R. BENNETT

Statistics of a synchronous binary message pulse train applied to a regenerative repeater are related to those of the original binary message, which is assumed ergodic. It is shown that the ensemble of possible message pulse trains is a nonstationary random process having a periodically varying mean and autocovariance. A spectral density is calculated which shows line spectral components at harmonics of the pulse rate and a continuous density function, both with intensity proportional to the square of the absolute value of the Fourier transform of the standard pulse at the frequency considered. The continuous component has many properties similar to thermal noise but differs in that, under certain conditions described, it can exhibit regularly spaced axis crossings, can be exactly predicted over finite intervals and is capable of producing discrete components when nonlinear operations are performed on it, even though no line spectral terms are originally present. The analytical results are applied to the problem of deriving a timing wave from the message pulse train by shock-exciting a tuned circuit with impulses occurring at the axis crossings.

I. INTRODUCTION

An ideal regenerative digital repeater is defined here as a nonlinear time-varying device with response expressed as the product of a staircase output vs. input function such as shown in Fig. 1(a) and a time-sampling function such as shown in Fig. 1(b). The first function (a) converts a continuous range of possible input values to a discrete or quantized set of output values. The second function (b) is a time-varying response function which ideally makes the repeater sensitive only during infinitesimally narrow time intervals with regular spacing $T = 1/f_r$. Taken together, the two functions quantize and sample the input wave at uniformly spaced instants of time. The input wave consists of a discrete-valued message component plus noise and distortion accumulated

in transmission from the previous repeater. If the noise and distortion are sufficiently small to confine the departure from the proper discrete message value within one quantization interval at the sampling instants, the regenerated wave matches the message and errorless transmission is achieved for any number of repeater spans.

To economize bandwidth, the actual pulses transmitted are not infinitesimally sharp, nor even rectangular as indicated in Fig. 1(b), but are multiples of a standard finite width pulse $g(t)$ which may be considered to be generated by a linear network from the sharp samples. A typical outgoing wave from the repeater is therefore expressible by:

$$x(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT), \qquad (1)$$

where the sequence a_n, n going from $-\infty$ to ∞, represents the message values. In much of our work we shall specialize our attention to the binary case, in which there are only two possible values of a_n, which we shall usually take as zero and unity. The corresponding output vs. input function of Fig. 1(a) then takes the form shown in Fig. 2. A typical outgoing wave train is shown in Fig. 3 for the message sequence

$$\cdots 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0 \cdots.$$

The ideal repeater requires an absolute clock to supply the timing control exemplified by Fig. 1(b). In practical transmission systems of considerable length, it may not be feasible to supply such absolute timing information. An alternate procedure, for example, derives the timing wave from the signal pulse train itself by applying a portion of the wave
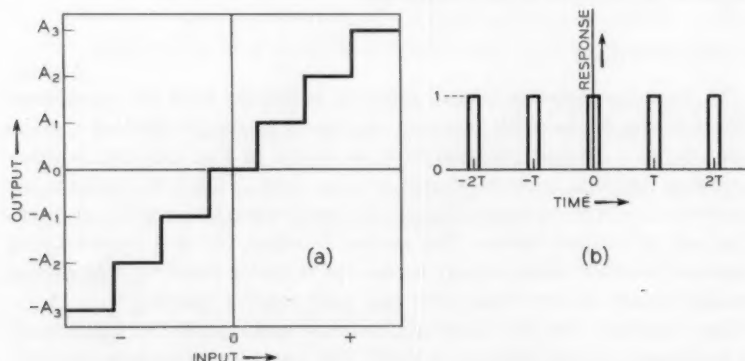

Fig. 1 — Specification of ideal regenerative repeater characteristic.

train to a circuit tuned to the pulse repetition frequency. If the circuit has a sufficiently high Q it continues to oscillate at the desired frequency during the gaps when no message pulses are present. Design of the tuned circuit requires a statistical analysis of the pulse train. This is one of many important statistical problems associated with practical, as distinguished from ideal, regenerative repeaters.

In this paper, we consider the following specific problems:

1. Properties of a digital message pulse train as a random noise source.

2. Derivation of the pulse repetition frequency from a pulse train by shock excitation of a tuned circuit. Under this topic, the following effects are studied:

    (a) message statistics

    (b) message pulse shape

    (c) Q

    (d) mistuning

    (e) noise.

3. Effect of time jitter in received pulse train on recovered analog signal.

The important problem of analyzing accumulated effects in a chain of nonideal regenerative repeaters is discussed in a companion paper by H. E. Rowe.[1] Both this paper and that of Rowe are intended to pro-
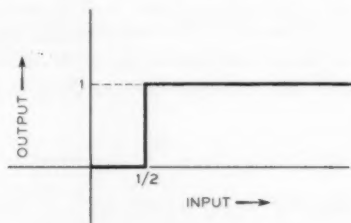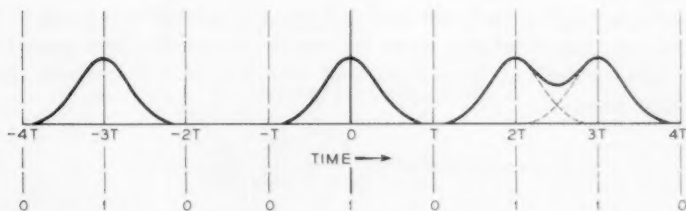

Fig. 2 — Binary output-vs.-input response.

Fig. 3 — Typical binary wave train.

vide analytical background for the two preceding papers in this issue.[2, 3] Reference is also made to a paper by E. D. Sunde[4] which treats many of the same problems in a somewhat different way and gives results for specific repeater embodiments.

II. A DIGITAL MESSAGE AS A RANDOM NOISE SOURCE

The digital message may be regarded as a random time series of discrete numbers. In the notation of stochastic processes, consider the ensemble of possible messages $M(t)$, where a typical $M(t)$ consists of the infinite discrete time sequence: $\cdots a_{-2}, a_{-1}, a_0, a_1, a_2 \cdots$. The values of the a's belong to a specified set of discrete numbers and each possible sequence has a probability of occurrence associated with it. We assume that the message ensemble is ergodic; that is, averages over the ensemble at fixed time are identical with averages over time in any member of the ensemble, except for a set of probability zero. It follows that the ensemble is also stationary; that is, averages over the ensemble do not depend on position in the sequence. For our purposes we do not go beyond second-order statistics and shall base our calculations on two ensemble averages, the ordinary mean and the autocovariance.

The mean m_1 is defined by the ensemble average for any n:

$$m_1 = \text{av } a_n . \tag{2}$$

The value of m_1 is a constant for the message ensemble. For example, in the binary case in which zero and one occur with equal probability m_1 is equal to one half. The autocovariance $R(n)$ is defined by the ensemble average

$$R(n) = \text{av } (a_k a_{k+n}) \quad \text{for fixed } k. \tag{3}$$

The value of $R(n)$ depends on n but not on k. It is determined by the dependency of successive message values. If the message consists of independent numbers, the value of $R(0)$ is the average squared message value and $R(n)$ is equal to the square of the mean for n not equal to zero, since the average of the product of independent quantities is equal to the product of their averages. From the ergodic property, these quantities may also be derived from almost all members of the ensemble individually, thus:

$$m_1 = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{n=-N}^{N} a_n , \tag{4}$$

$$R(n) = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{k=-N}^{N} a_k a_{k+n} . \tag{5}$$

The actual wave sent over the line from one repeater to the next is the function $x(t)$ given by (1). We may regard $x(t)$ as a typical member of the pulse ensemble $\{x(t)\}$. Unlike the message ensemble $\{M(t)\}$, the pulse ensemble $\{x(t)\}$ consists of continuous functions of time rather than discrete time series. We shall now relate the statistics of the pulse ensemble to those of the discrete message ensemble.

We first evaluate the ensemble average of $\{x(t)\}$ by holding t fixed and averaging over all members. Then, since the average of a sum is equal to the sum of individual averages, and since the standard pulse $g(t)$ is the same for all members of the ensemble, we find

$$
\begin{aligned}
\operatorname{av} \{x(t)\} &= \operatorname{av} \sum_{n=-\infty}^{\infty} a_n g(t - nT) \\
&= \sum_{n=-\infty}^{\infty} \operatorname{av}(a_n) g(t - nT) \qquad (6)\\
&= m_1 \sum_{n=-\infty}^{\infty} g(t - nT).
\end{aligned}
$$

It thus appears that the ensemble average varies with time. This is not really unexpected since it seems reasonable that we should find values in the middle of a pulse interval to be distributed quite differently from those at the beginning or end. In fact, we should expect the ensemble average to vary periodically at the pulsing rate, and we readily demonstrate this by noting that

$$
\begin{aligned}
\operatorname{av} \{x(t + T)\} &= m_1 \sum_{n=-\infty}^{\infty} g(t + T - nT) \\
&= m_1 \sum_{n=-\infty}^{\infty} g[t - (n - 1)T] = \operatorname{av} x(t)
\end{aligned}
\qquad (7)
$$

by substitution of the summation index $n' = n - 1$.

The periodicity of the ensemble average suggests a formal Fourier series expansion, thus

$$
\operatorname{av} x(t) = \sum_{m=-\infty}^{\infty} c_m \exp (2m\pi j f_r t), \qquad (8)
$$

$$
\begin{aligned}
c_m &= f_r \int_0^T \operatorname{av} x(t) \exp (-2m\pi j f_r t) \, dt \\
&= m_1 f_r \sum_{n=-\infty}^{\infty} \int_0^T g(t - nT) \exp (-2m\pi j f_r t) \, dt.
\end{aligned}
\qquad (9)
$$

We now substitute $t - nT = u$ in the integral. The limits then become $-nT$ to $-(n - 1)T$ while the exponential becomes $\exp (-2m\pi j f_r u)$,

since $f_r T = 1$. The summation of finite integrals with adjoining limits may then be replaced by a single infinite integral, giving

$$c_m = m_1 f_r \int_{-\infty}^{\infty} g(u) \exp\left(-2m\pi j f_r u\right) du. \tag{10}$$

If we introduce the Fourier transform of the standard unit pulse by

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp\left(-2\pi j f t\right) dt \tag{11}$$

we observe that

$$c_m = m_1 f_r G(m f_r), \tag{12}$$

and, hence,

$$\text{av } \{x(t)\} = m_1 f_r \sum_{m=-\infty}^{\infty} G(m f_r) \exp\left(2m\pi j f_r t\right). \tag{13}$$

The ensemble average is thus expressed as a Fourier series in time with the amplitude of the mth harmonic of the signaling frequency proportional to the amplitude of the spectral representation of the unit pulse at that same harmonic frequency. We note that an ensemble consisting of precisely this set of harmonics and no other terms would exhibit just this same result for its ensemble average. It is convenient, therefore, to resolve our pulse ensemble into a set of such discrete nonrandom harmonic components plus a remainder which has random properties and zero mean. We consider therefore the ensemble

$$\{y(t)\} = \{x(t)\} - \text{av } \{x(t)\} = \sum_{n=-\infty}^{\infty} (a_n - m_1)g(t - nT). \tag{14}$$

Then,

$$\text{av } \{y(t)\} = 0. \tag{15}$$

We explore farther by evaluating the autocovariance function $R_y(\tau, t)$ for the ensemble $\{y(t)\}$. By definition,

$$R_y(\tau, t)$$

$$= \text{av } \{y(t)y(t + \tau)\}$$

$$= \text{av } \left\{ \sum_{m=-\infty}^{\infty} (a_m - m_1)g(t - mT) \sum_{n=-\infty}^{\infty} (a_n - m_1)g(t - nT + \tau) \right\} \tag{16}$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \text{av}[(a_m - m_1)(a_n - m_1)]g(t - mT)g(t - nT + \tau)$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} [R(m - n) - m_1^2]g(t - mT)g(t - nT + \tau).$$

By replacing t by $t + T$ and increasing both summation indices by unity, we see that $R_y(\tau, t)$ is periodic in t with period T. Hence, as in the case of the mean,

$$R_y(\tau, t) = \sum_{k=-\infty}^{\infty} d_k \exp (2k\pi jf_r t), \tag{17}$$

$$d_k = f_r \int_0^T R_y(\tau, t) \exp (-2k\pi jf_r t) \, dt$$

$$= f_r \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} [R(m - n) - m_1^2] \tag{18}$$

$$\int_0^T g(t - mT)g(t - nT + \tau) \exp (-2k\pi jf_r t) \, dt.$$

We substitute $m - n = m'$, $t - nT = t'$, and find after some rearrangement analogous to that used in obtaining (10):

$$d_k = f_r \sum_{m=-\infty}^{\infty} [R(m) - m_1^2]F(\tau + mT, kf_r), \tag{19}$$

where

$$F(\tau, f) = \int_{-\infty}^{\infty} g(t)g(t + \tau) \exp (-j2\pi ft) \, dt. \tag{20}$$

By the convolution theorem,

$$F(\tau, f) = \int_{-\infty}^{\infty} G(\lambda)G^*(\lambda + f) \exp [-j2\pi\tau(\lambda + f)] \, d\lambda. \tag{21}$$

In the special case of independent message values,

$$R(0) = \operatorname{av} (a_n^2) = m_2; \qquad R(n) = \operatorname{av}^2 (a_n) = m_1^2, n \neq 0 \tag{22}$$

and

$$d_k = f_r \operatorname{var} (a_n) F(\tau, kf_r), \tag{23}$$

where $\operatorname{var} (a_n)$ is the variance of the a's defined by

$$\operatorname{var} (a_n) = m_2 - m_1^2 = \operatorname{av} (a_n^2) - \operatorname{av}^2 a_n. \tag{24}$$

There is no reason, in general, for $F(\tau, kf_r)$ to vanish for all nonzero values of k, and hence the autocovariance of the y-ensemble does not reduce to an expression independent of time. It appears, therefore, that separating out the periodic components which account for the fluctuating mean does not make the remainder a stationary process. We therefore cannot invoke the Fourier transform relationship between autocovariance and power spectrum to find a power spectrum for our pulse ensemble, even after the periodic components have been removed.

A procedure which has often been used in similar situations is to make the process stationary by assuming a random phase relationship between the members of the ensemble. We do not wish to do this here because synchronous phase is an important property to be preserved in digital regeneration. However, there is a meaningful definition of a power spectrum which can be used without reference to the fluctuating auto-covariance. We shall calculate the spectral density in this way, but, as might be expected, the resulting function is the same as would be obtained by randomizing the phases of the ensemble.

Consider first the ensemble $y_N(t)$ which includes only the pulses from $n = -N$ to $n = N$:

$$y_N(t) = \sum_{n=-N}^{N} (a_n - m_1)g(t - nT). \tag{25}$$

Then, for unit pulses which possess a Fourier transform, the Fourier transform $S_N(f)$ of $y_N(t)$ exists and is given by

$$S_N(f) = \sum_{n=-N}^{N} (a_n - m_1)g(f) \exp(-2n\pi jfT). \tag{26}$$

By Parseval's theorem,

$$\int_{-\infty}^{\infty} y_N^2(t) \, dt = \int_{-\infty}^{\infty} |S_N(f)|^2 \, df. \tag{27}$$

Let the average of the ensemble $y_N^2(t)$ over the interval $-NT$ to NT be represented by $\mathrm{av}_N y_N^2(t)$. Then,

$$\begin{aligned}
\mathrm{av}_N \, y_N^2(t) &= \frac{1}{(2N+1)T} \int_{-NT}^{NT} \mathrm{av} \, y_N^2(t) \, dt \\
&= \frac{1}{(2N+1)T} \int_{-\infty}^{\infty} \mathrm{av} \, |S_N(f)|^2 \, df \\
&= \frac{1}{(2N+1)T} \left[\int_{NT}^{\infty} + \int_{-\infty}^{NT} \right] \mathrm{av} \, y_N^2(t) \, dt.
\end{aligned} \tag{28}$$

Also,

$$\begin{aligned}
\mathrm{av}_\infty \, y_\infty^2(t) &= \lim_{N\to\infty} \mathrm{av}_N \, y_N^2(t) \\
&= \lim_{N\to\infty} \int_{-\infty}^{\infty} \frac{\mathrm{av} \, |S_N(f)|^2}{(2N+1)T} \, df \\
&= \int_{-\infty}^{\infty} w(f) \, df,
\end{aligned} \tag{29}$$

where

$$w(f) = \lim_{N\to\infty} \frac{\mathrm{av} \, |S_N(f)|^2}{(2N+1)T}. \tag{30}$$

From (29), $w(f)$ as defined by (30) satisfies the requirement for a spectral density. From (26),

$$\text{av} \mid S_N(f) \mid^2$$

$$= \text{av} \left[S_N(f) S_N{}^*(f) \right]$$

$$= \sum_{m=-N}^{N} \sum_{n=-N}^{N} \text{av} \left[(a_n - m_1)(a_m - m_1) \right] G(f) G^*(f) \exp \left[2\pi jfT(m - n) \right] \quad (31)$$

$$= \sum_{m=-N}^{N} \sum_{n=-N}^{N} \left[R(m - n) - m_1{}^2 \right] \mid G(f) \mid^2 \exp \left[2\pi jfT(m - n) \right].$$

Let $m - n = k$, giving

$$\text{av} \mid S_N(f) \mid^2 = \sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} \left[R(k) - m_1{}^2 \right] \mid G(f) \mid^2 \exp \left(2\pi jkfT \right). \quad (32)$$

The order of summation can be interchanged by the formula:

$$\sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} = \sum_{k=-2N}^{0} \sum_{m=-N}^{k+N} + \sum_{k=1}^{2N} \sum_{m=k-N}^{N}. \quad (33)$$

Since the summand does not depend on m, we perform the summation on m by counting the terms. The result is:

$$\text{av} \mid S_N(f) \mid^2 = (2N + 1) \mid G(f) \mid^2 \left\{ R(0) - m_1{}^2 \right.$$

$$\left. + 2 \sum_{k=1}^{2N} \left(1 - \frac{k}{2N + 1} \right) \left[R(k) - m_1{}^2 \right] \cos 2\pi kfT \right\}. \quad (34)$$

From (30), then,

$$w(f) \doteq f_r \mid G(f) \mid^2 \left\{ R(0) - m_1{}^2 + 2 \sum_{k=1}^{\infty} \left[R(k) - m_1{}^2 \right] \cos 2\pi kfT \right\}. \quad (35)$$

In the special case of independent message values, $R(0) = m_2 = m_1$ and $R(k) = m_1{}^2$ for $k \neq 0$, giving a result in agreement with the well-known solution[5] based on a random phase ensemble average, namely

$$w(f) = f_r \, m_1(1 - m_1) \mid G(f) \mid^2. \quad (36)$$

This spectral density function is defined for both positive and negative frequencies. If the densities at negative frequencies are added to those of the corresponding positive frequencies, the above result is multiplied by two.

The spectrum of $\{y(t)\}$ is thus continuous and is proportional to the squared absolute value of the Fourier transform of the unit pulse. The spectrum of the actual pulse ensemble has added to the above con-

tinuous spectrum a set of line spectral components given by (13). The line spectral terms occur at harmonics of the signaling frequency and have amplitudes proportional to the unit pulse spectrum evaluated at the harmonic frequencies. If the pulse spectrum vanishes at any harmonic that harmonic does not appear in the $\{x(t)\}$ ensemble spectrum. Equation (13) can also be written in the form

$$
\begin{aligned}
\text{av} \{x(t)\} = {} & m_1 f_r G(0) \\
& + 2m_1 f_r \sum_{m=1}^{\infty} \mid G(m f_r) \mid \cos [m\omega_r t + \text{ph } G(m f_r)].
\end{aligned}
\tag{37}
$$

From this expression it is clear that the dc component has mean square $m_1^2 f_r^2 G^2(0)$, while the mean square value of the mth harmonic is

$$
2m_1^2 f_r^2 \mid G(m f_r) \mid^2.
$$

Fig. 4 gives curves of the continuous and line spectral density components for independent binary on-off signaling with various standard pulse shapes. For this case, if p_1 is the probability of the value unity, it follows that

$$
m_2 = p_1 ; \qquad m_2 - m_1^2 = p_1(1 - p_1).
\tag{38}
$$

The continuous part of the pulse train spectrum has much in common with thermal noise. An audible band selected between adjacent harmonics sounds to the ear like thermal noise. If the signaling rate is placed well above the noise band a random on-off pulse train serves adequately as a hiss source in a vocoder synthesizer. Oscillograms and spectrograms of narrow-band telegraph noise not including a harmonic would appear to the eye to be indistinguishable from thermal noise. The curves of Fig. 4 provide useful quantitative data relating the amount of noise produced in this way for specific cases.

It must not be concluded, however, that random telegraph noise is in all respects equivalent to thermal noise with the same spectral density. The telegraph noise remains a nonstationary process and has a phase structure among its components which thermal noise does not possess. The nonstationary properties of the telegraph wave ensemble are not the most general but are of a special kind which have sometimes been described as "periodically stationary." We have suggested the name "cyclostationary" for this type of process, i.e., one in which the ensemble statistics vary periodically with time. Analogously, the name "cyclo-ergodic" will be used for cyclostationary processes in which statistics over the ensemble for fixed time are the same as statistics over the in-
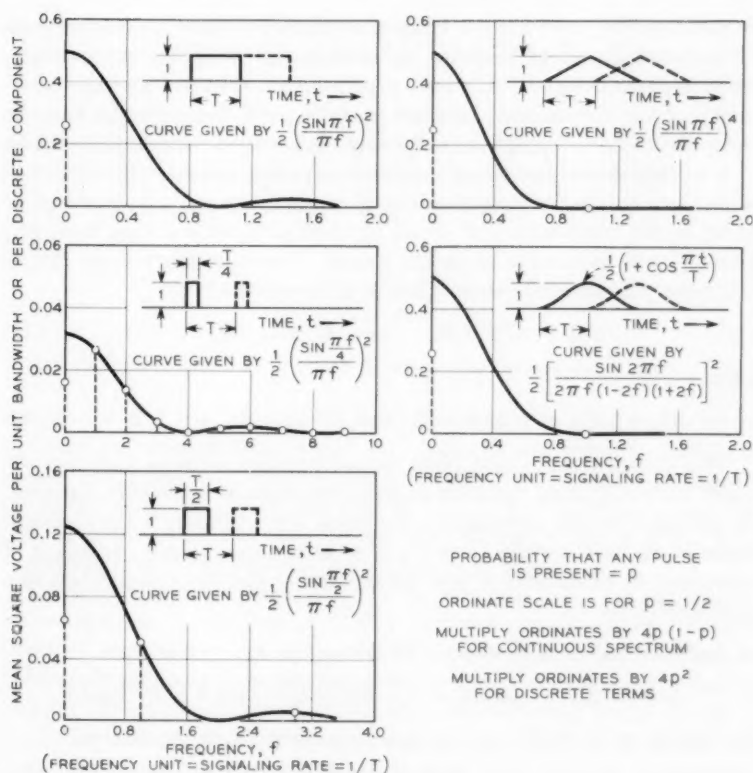
Fig. 4 — Continuous and line spectral density components for independent binary on-off signaling with specific pulse shapes.

stants of time differing from the fixed instant by multiples of the period in almost all individual members of the ensemble. A cycloergodic ensemble is cyclostationary but the converse is not necessarily true.

An important practical difference between telegraph noise and thermal noise is that the former can be generated deterministically if the message values are known or can be found. This follows because the wave form is completely determined by the sequence of message values and the standard unit pulse wave form. Use of telegraph noise for masking other waves is therefore subject to the qualification that, if the telegraph message can be read, the noise can be removed. Reading the message becomes more and more difficult as the ratio of bandwidth to signaling frequency

is made smaller, and the overlapping of adjacent pulses in time becomes correspondingly greater. From the standpoint of digital transmission, reading the message is, of course, a paramount objective and the bandwidth of the transmission medium must be made large enough and the noise and distortion small enough to enable a correct reading to be made.

A further salient feature of the telegraph noise ensemble as contrasted with thermal noise is the possibility of regularly spaced axis crossings or zeros of the received wave. From (1) and (14) we see that both $x(t)$ and $y(t)$ will have zeros at points spaced T seconds apart if the standard pulse $g(t)$ has such zeros. That is, if for some t_0

$$g(t_0 + nT) = 0, \qquad n = 0, \pm 1, \pm 2, \cdots , \tag{39}$$

then

$$x(t_0 + nT) = y(t_0 + nT) = 0, \qquad n = 0, \pm 1, \pm 2, \cdots . \tag{40}$$

This does not mean that there cannot be other zeros as well, and the other zeros in general can have irregular spacings. It is possible, however, to exclude irregular zeros in specific cases. For example, if the standard pulse is the unit impulse response of a series-tuned circuit consisting of resistance R, inductance L, and capacitance C, the function $g(t)$ becomes $g_0(t)$ defined by

$$\omega_0 L g_0(t) = (\omega_0^2 + \alpha^2)^{1/2} \exp(-\alpha t) \cos(\omega_0 t + \theta); \qquad t > 0$$
$$\alpha = R/2L; \qquad \omega_0^2 L C = 1 - R^2 C/4L; \qquad \tan \theta = R/(2\omega_0 L). \tag{41}$$

By setting $\omega_0 = 2\pi/T$, we obtain nulls at t_0 and t_1 in the interval

$$0 \leqq t \leqq T,$$

where

$$2\pi t_0/T = \pi/2 - \theta,$$
$$2\pi t_1/T = 3\pi/2 - \theta. \tag{42}$$

These represent the only nulls in the interval from zero to T and they repeat at $t_0 + nT$ and $t_1 + nT$ for all integer values of n, as shown in the full-line waveform of Fig. 5. Furthermore, $g_0(t)$ changes in sign from positive to negative at $t = t_0 + nT$ and from negative to positive at $t = t_1 + nT$. No other sign changes occur. If we consider $g_0(t - mT)$ with m any positive or negative integer, we note that the values between $t_0 + nT$ and $t_1 + nT$ are the same as those of $g_0(t)$ between $t_0 + (n - m)T$ and $t_1 + (n - m)T$, and hence they are all negative. Likewise, the values of $g_0(t - mT)$ between $t_1 + nT$ and $t_0 + (n + 1)T$ are all positive. The
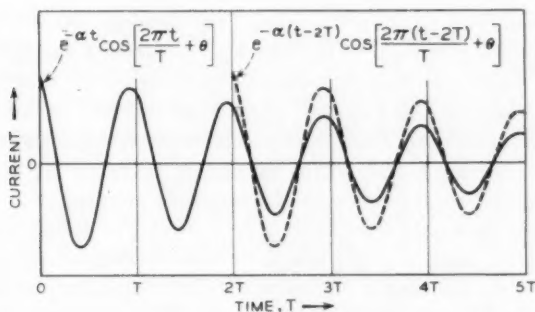
Fig. 5 — Response of tuned circuit to synchronous impulses.

dashed curve of Fig. 5 shows the relation for $m = 2$. The conclusion is that the summations from which $x(t)$ and $y(t)$ are formed consist of waves having common nulls and the same sign between nulls. It follows that both $x(t)$ and $y(t)$ for the case of $g(t) - g_0(t)$ have nulls only at $t_0 + nT$ and $t_1 + nT$, the axis crossings are from plus to minus at $t_0 + nT$, and the axis crossings are from minus to plus at $t_1 + nT$.

We have thus proved that, if $g_0(t)$ is the standard pulse, the continuous noise spectrum arising from a random choice of signaling pulses has no effect on the nulls of the composite wave. Telegraph noise therefore is profoundly different from thermal noise in that it does not perturb axis crossings in a situation in which thermal noise certainly would. Production of an invariant null spacing permits the recovery of the timing or clock wave from the pulse train and is accordingly of vital importance in the self-timed regenerative repeater.

In general, the standard pulse will not have the invariant null property exhibited by the special function $g_0(t)$. It may, however, be possible to convert the actual $g(t)$ into a new pulse which does have the required property. We note that the Fourier transform of $g_0(t)$ is given by

$$G_0(f) = \frac{j2\pi fC}{1 + j2\pi fRC - 4\pi^2 f^2 LC}. \tag{43}$$

An ensemble with standard pulse $g(t)$ and transform $G(f)$ can be converted into one with standard pulse $g_0(t - T_0)$ by inserting a linear network with transmittance function

$$Y(f) = \frac{G_0(f)}{G(f)} \exp(-j2\pi fT_0), \tag{44}$$

provided that $Y(f)$ is physically realizable. Since $G_0(f)$ is the response of

a tuned circuit to a constant spectrum, the problem reduces to the re-
alizability of the reciprocal of $G(f)$ with an arbitrary delay factor.

Suppose, however, that the received pulse shape is such that $Y(f)$
can not be realized. For example, (3) for the discrete or line spectrum
components shows that, if the pulse transform $G(f)$ vanishes at the pulse
repetition frequency f_r and all its harmonics, there are no discrete fre-
quency components representing the signaling frequency in the pulse
train. In (44) we would then have $G(f) = 0$ at $f = f_r$ and, since $F_0(f_r)$
is not zero, the value of $Y(f_r)$ would have to be infinite. It might seem
that in this case the timing information is lost. However, we shall show
that this difficulty only exists for linear time-invariant methods of deriv-
ing the timing information.

To see how we can materialize a discrete component from a con-
tinuous spectrum, consider a specific case of an uncurbed rectangular
signaling pulse

$$g(t) = 1; \qquad 0 < t < T. \tag{45}$$

Then, from (11),

$$G(f) = (\pi f)^{-1} \exp(-j\pi f/f_r) \sin(\pi f/f_r). \tag{46}$$

It is clear that $G(nf_r)$ vanishes for all integer values of n. If a pulse by
itself contains zero amplitude at all harmonics of f_r, no succession of
pulses can have nonzero amplitude at these frequencies. However, we
note that, in a binary on-off system using these uncurbed pulses, a wave
such as that shown in Fig. 6 is obtained, in which every message digram
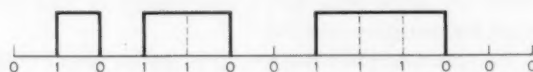

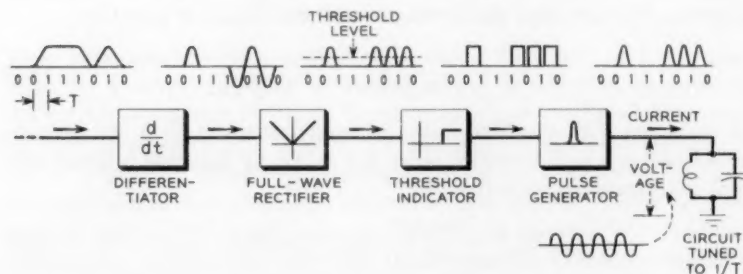Fig. 6 — Train of uncurbed rectangular pulses.

Fig. 7 — Recovery of timing wave from transitions of binary pulse train.

01 produces an upward transition, every digram 10 produces a downward transition, and the wave is constant for unbroken sequences of 1's or 0's. If this wave is applied to the circuit of Fig. 7, which includes a differentiator, full-wave rectifier, threshold triggered pulse generator and tuned circuit, an output wave containing pulses of form

$$g_0(t - nT + T_0)$$

is obtained where n represents the integer values in the message sequence at which 01 or 10 digrams occur. The effective message source is transformed from the original sequence of zero and unity values to a new sequence in which the unity values are associated with 01 or 10 transitions and the zeros with 11 or 00 digrams. A new continuous spectrum is thereby obtained and a discrete line spectrum absent from the original wave is created. Nonlinear operations have been used to attain this end. The system will still operate if either (but not both) the rectifier or the nonlinear triggered pulse generator is omitted. If we were dealing with thermal noise, even nonlinear operations would not suffice to construct a discrete spectrum.

The impulse response of a tuned circuit is only one example of a standard pulse shape which yields correct timing information from a synchronous telegraph pulse train. Another standard pulse having the same essential properties is the response of the same tuned circuit to a rectangular pulse of duration equal to half the signaling period. When the exciting pulse has unit height, the response becomes $g_1(t)$, defined by

$$\omega_0 L g_1(t) = \exp(-\alpha t) \begin{cases} \sin \omega_0 t; & 0 < t < T/2 \\ \sin \omega_0 t - \exp(\alpha T/2) \sin \omega_0(t - T/2); & t > T/2. \end{cases} \tag{47}$$

When $\omega_0 T = 2\pi$, it may be verified that $g_1(t)$ has positive-going axis crossings at $t = 0, T, 2T, \cdots$; negative going axis crossings at $t = T/2$, $3T/2, 5T/2, \cdots$; and no other axis crossings. Hence the summation of any such pulses beginning only at multiples of T has the same nulls and no others.

The problem of producing regularly spaced axis crossings from a received pulse train is studied in more detail in the next section, which treats the problems associated with derivation of timing information. We conclude our discussion here with some further remarks about secondary pulse trains triggered from transitions in a primary pulse train. Such secondary trains are of major importance in pulse frequency derivation techniques. The example given above of a primary uncurbed rectangular pulse train is not of much interest in digital transmission,

since it would require excessive bandwidth to approximate the square waveforms. A more practical type of pulse is the raised cosine of Fig. 8. Such a waveform can be transmitted with fair accuracy over a bandwidth slightly less than the signaling frequency. Closest spacing without overlapping is obtained when one such pulse is started at the instant a preceding one reaches its peak. A flat-topped resultant wave then occurs between successive individual pulse peaks, as shown in Fig. 8. A train of raised cosine pulses can therefore give timing information only on the transitions 01 and 10, and the circuit of Fig. 7 is appropriate for signaling frequency derivation. It may be verified from Fig. 4 that the raised cosine pulse has null values in its Fourier transform at the signaling frequency and its harmonics.

The message associated with the transition pulse train is, of course, different from the original message and has a different set of statistics. It is of interest to calculate the relation between the two for the case of binary on-off message source with independent signal values. For such a
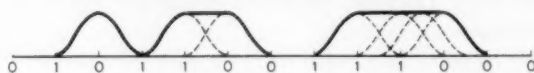

Fig. 8 — Train of raised cosine pulses.

source, the values of a_n are selected independently, with the probability p_1 that unity occurs and $1 - p_1$ that zero occurs in any position. Then the mean first and second powers are given by

$$m_1 = m_2 = p_1 \tag{48}$$

and the autocovariance by

$$R(0) = m_1, \qquad R(n) = m_1^2, \qquad n \neq 0. \tag{49}$$

The rule for constructing the transition message is given by the following table:

Original Message	Transition Message
1 preceded by 0	1
1 preceded by 1	0
0 preceded by 0	0
0 preceded by 1	1

The value 1 occurs in the transition message from two events each having probability $m_1(1 - m_1)$ and hence has probability $2m_1(1 - m_1)$. The

mean first and second powers of the transition series are therefore given by

$$r_1 = r_2 = 2m_1(1 - m_1). \tag{50}$$

We shall represent the autocovariance of the transition message ensemble by $\rho(n)$. The value of $\rho(0)$ is simply the mean square, and hence

$$\rho(0) = 2m_1(1 - m_1) = r_1. \tag{51}$$

To evaluate $\rho(1)$ we note that the product of adjacent values in the transition message is zero except when the original values are 10 preceded by 0, or 01 preceded by 1. The probability of obtaining the first is $m_1(1 - m_1)^2$ and of obtaining the second $m_1^2(1 - m_1)$. Hence the probability of a 11 sequence in the transition messages is given by the sum of these two probabilities, and since we obtain unity for the product in this case and zero otherwise,

$$\rho(1) = m_1(1 - m_1)^2 + m_1^2(1 - m_1) = m_1(1 - m_1) = r_1/2. \tag{52}$$

For values of n greater than one the constraint imposed by the transitions effectively disappears, since we may fill in all possible 0 and 1 values between the critical end points. Thus, in the case of $n = 2$ we obtain values of unity in the transition series two intervals apart for the following original message sequences:

$$
\begin{array}{cccc}
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0.
\end{array}
$$

These have total probability

$$\rho(2) = 4m_1^2(1 - m_1)^2 = r_1^2. \tag{53}$$

To evaluate $\rho(3)$, we merely fill in all possible 1's and 0's in a column between the second and third above and obtain the same total probability because the compound event represented by the first and last two columns is independent of that represented by the middle column. The probability of the first compound event is r_1^2, as calculated above, and that of a second is unity. Hence the complete autocovariance is expressed by

$$
\begin{aligned}
\rho(0) &= r_1, \\
\rho(1) &= \rho(-1) = r_1/2, \qquad \text{for} \quad |n| > 1. \tag{54} \\
\rho(n) &= \rho(-n) = r_1^2
\end{aligned}
$$

Thus, the transition series formed from the independent binary series becomes a digram series; that is, one in which adjacent values are dependent but nonadjacent values are independent. The discrete components in the spectrum are the same as in the independent case except that m_1 is replaced by r_1. The continuous spectrum is evaluated by substituting ρ for R and r_1 for m_1 in (35). The result after r_1 has been replaced by its value from (50) is

$$w(f) =$$

$$2m_1(1 - m_1) f_r \mid G(f) \mid^2 [1 - 2m_1 + 2m_1^2 + (1 - 2m_1)^2 \cos 2\pi fT]. \quad (55)$$

In the special case in which $m_1 = \frac{1}{2}$, this reduces to the same spectral density obtained for the independent message case, (36).

III. DERIVATION OF TIMING INFORMATION FROM A RANDOM TELEGRAPH MESSAGE

The existence of standard telegraph pulses which shock-excite a tuned circuit to produce a sustained oscillation at the signaling rate for almost all message sequences has been demonstrated by example in Section II.

In practical timing recovery circuits the ideal of a constant-amplitude correct-frequency output is not perfectly realizable. Factors which influence the result include the message pulse pattern, the Q of the tuned circuit, the presence of noise and the precision to which the natural oscillation frequency of the tuned circuit can be made to match the signaling frequency.

We shall analyze the problem of timing recovery under the conditions in which the ideal performance is almost, but not quite, obtained. The errors in this case are relatively small in magnitude and corresponding approximations can be made in the calculations which will show with sufficient accuracy what the quantitative requirements must be to hold the errors within specified small ranges.

Assume that the tuned circuit oscillates at a frequency very near to f_r, the signaling frequency, and that Q is sufficiently high to make the influence of any one pulse on the amplitude of oscillation slight. Consider the ensemble of possible responses in the interval from $t = 0$ to $t = T$, with the process having begun at $t = -\infty$. Almost all the waveforms are approximately sinusoidal and the negative-going axis crossings cluster about $t = t_c$, where t_c is defined by

$$\text{av } x(t_c) = 0; \qquad \text{av } x'(t_c) < 0. \quad (56)$$

The actual axis crossing for the typical member of the ensemble occurs

at $t_c + \epsilon/2\pi f_r$, where ϵ is a small angle representing the phase error and has a distribution of values over the ensemble. We assume that the waveform through this axis crossing is that of a sine wave of frequency f_r and amplitude equal to the value of $x(t)$ one-quarter period after the axis crossing, i.e.:

$$x(t) \cong -x(t_c + \epsilon/\omega_r + T/4) \sin [\omega_r(t - t_c) - \epsilon], \qquad (57)$$

with $| \epsilon | \ll 1$. Then, retaining only first powers of ϵ, we have

$$x(t_c) \doteq \epsilon \, x(t_c + T/4). \qquad (58)$$

We further assume that $x(t_c + T/4)$ can be written in the form

$$x(t_c + T/4) = (1 + \zeta) \text{ av } x(t_c + T/4), \qquad (59)$$

where $| \zeta | \ll 1$. By averaging both sides of this equation over the ensemble we deduce that av $\zeta = 0$.

It follows from the above assumptions that

$$
\begin{aligned}
\text{av } x(t_c) &\doteq \text{av } [\epsilon(1 + \zeta) \text{ av } x(t_c + T/4)] \\
&\doteq \text{av } \epsilon \text{ av } x(t_c + T/4).
\end{aligned}
\qquad (60)
$$

Hence, av $x(t_c) = 0$ implies av $\epsilon \doteq 0$. The principal quantity of interest is the rms phase error, which is the square root of av ϵ^2. By the assumed cycloergodicity of the process, an ensemble average of ϵ^2 is equal to an average of the value of ϵ^2 over all intervals of any member of the ensemble. From (58) and (59),

$$
\begin{aligned}
\text{av } x^2(t_c) &\doteq [\epsilon^2(1 + 2\zeta + \zeta^2) \text{ av}^2 \, x(t_c + T/4)] \\
&\doteq \text{av } \epsilon^2 \text{ av}^2 \, x(t_c + T/4).
\end{aligned}
\qquad (61)
$$

Hence, to the first order of small quantities,

$$\text{av } \epsilon^2 = \frac{\text{av } x^2(t_c)}{\text{av}^2 \, x(t_c + T/4)}. \qquad (62)$$

The mean square phase error is thus expressed in terms calculable from the general expressions previously derived for means and covariances of the pulse ensemble.

It is also of interest to evaluate the mean square value of ζ to validate the treatment of ζ as a small quantity and to estimate the amount of amplitude variation. From (59),

$$
\begin{aligned}
\text{av } x^2(t_c + T/4) &= \text{av } (1 + 2\zeta + \zeta^2) \text{ av}^2 \, x(t_c + T/4) \\
&= (1 + \text{av } \zeta^2) \text{ av}^2 \, x(t_c + T/4).
\end{aligned}
\qquad (63)
$$

Hence,

$$\text{av } \zeta^2 = \frac{\text{av } y^2(t_c + T/4)}{\text{av}^2 \, x(t_c + T/4)} = \frac{R_y(0, t_c + T/4)}{\text{av}^2 \, x(t_c + T/4)}. \tag{64}$$

It is to be noted that the above treatment evaluates only the mean square total phase error and does not give a frequency resolution. The spectral composition is important, particularly in the case of a chain of repeaters and is discussed in the companion paper by Rowe.[1]

We next calculate the average values appearing in the equations for mean square phase error and mean square amplitude ratio variation for specific pulse shapes. We assume that the circuit is not tuned to the pulse repetition frequency so that a basis for estimating the error from mistuning can be established. For the case in which $g(t) = g_0(t)$, the unit impulse response of a single tuned circuit, we write (41) in the form

$$g_0(t) = A \, \exp \, (-\alpha t) \cos \, (\omega_0 t + \theta); \qquad t > 0, \tag{65}$$

where

$$A = \omega_0 (4Q^2 + 1)^{1/2}/2RQ^2; \qquad \alpha = \omega_0/2Q;$$
$$Q = \omega_0 L/R, \qquad \tan \theta = 1/2Q. \tag{66}$$

From (6), the ensemble average of $x(t)$ in the interval $0 \leqq t \leqq T$ is

$$\text{av } \{x(t)\} = m_1 \sum_{n=-\infty}^{\infty} g_0(t - nT)$$
$$= m_1 A \sum_{n=-\infty}^{0} \text{Re} \exp \, [(j - \alpha)\omega_0(t - nT) + j\theta]. \tag{67}$$

The series is geometric and may be summed to obtain

$$\text{av } x(t) = \frac{m_1 A \, \exp \, [-\omega_0(t - T/2)/2Q]}{2RQ^2(\cosh^2 \omega_0 T/4Q - \cos^2 \omega_0 T/2)}$$

$$\cdot \left\{ \sinh \frac{\omega_0 T}{4Q} \cos \frac{\omega_0 T}{2} \cos \left[\omega_0 \left(t - \frac{T}{2} \right) + \theta \right] \right. \tag{68}$$

$$\left. - \cosh \frac{\omega_0 T}{4Q} \sin \frac{\omega_0 T}{2} \sin \, [\omega_0(t - T/2) + \theta] \right\}.$$

The value of t_c in the noise-free mistuned case can be obtained readily from (68) by equating the terms inclosed in the braces to zero, giving

$$t_c = T/2 - \theta/\omega_0 + \omega_0^{-1} \text{ arc tan } [\tanh \, (\omega_0 T/4Q) \cot \, (\omega_0 T/2)]. \tag{69}$$

We note that t_c approaches $T/4$ as Q approaches infinity. If we assume perfect tuning by setting $\omega_0 T = 2\pi$, (68) becomes

$$\text{av } x(t) =$$

$$\frac{m_1\omega_r(4Q^2 + 1)^{1/2} \exp\left[-(\omega_r t - \pi)/2Q\right] \cos\left[\omega_r(t - T/2) + \theta\right]}{4RQ^2 \sinh \pi/2Q}. \tag{70}$$

That is, the average waveform of the tuned circuit response becomes a damped sine wave of the correct frequency. When Q is large, the damping is small. The limiting form as Q approaches infinity is an undamped sine wave:

$$\lim_{Q \to \infty} \text{av } x(t) = \frac{m_1\omega_r}{\pi R} \cos \omega_r(t - T/2). \tag{71}$$

This result is perfect for frequency recovery. It is to be noted that it does not matter whether the message values are independent or not. In the special case in which pulses are generated by transitions only, m_1 would be replaced by $r_1 = 2m_1 (1 - m_1)$, as given by (50).

To complete the estimate of rms phase error and amplitude ratio variation, we must evaluate the average value of $x(t)$ one-quarter period away from t_c, and the average squares of $x(t)$ at $t = t_c$ and $t = t_c + T/4$. The value of av $x(t_c + T/4)$ is obtained by substituting $t = t_c + T/4$ in (58). The evaluation of average squares may be done in general from the covariance formula of either (16) or (17)–(21), recalling that

$$\text{av } x^2(t) = \text{av}^2 x(t) + \text{av } y^2(t), \tag{72}$$

$$\text{av } y^2(t) = R_y(0, t). \tag{73}$$

Two binary cases of considerable importance which simplify considerably are those of completely independent message values and the message derived from transitions between independent values. These are:

Independent Binary Message

$$R(0) = m_1; \qquad R(n) = m_1^2; \qquad n \neq 0,$$

$$R_y(\tau, t) = m_1(1 - m_1) \sum_{m=-\infty}^{\infty} g(t - mT)g(t - mT + \tau), \tag{74}$$

$$\text{av } y^2(t) = R_y(0, t) = m_1(1 - m_1) \sum_{m=-\infty}^{\infty} g^2(t - mT).$$

Transition Message for Independent Binary Values

$$R(0) = r_1 = 2m_1(1 - m_1), R(1) = R(-1)$$
$$= r_1/2, R(n) = r_1^2; |n| > 1, \tag{75}$$

$$R_y(\tau, t) = r_1 \sum_{m=-\infty}^{\infty} g(t - mT)(1 - r_1)g(t - mT + \tau)$$
$$+ (\tfrac{1}{2} - r_1)g[t - (m - 1)T + \tau] + (\tfrac{1}{2} - r_1)g[t - (m + 1)T + \tau], \tag{76}$$

$$\text{av } y^2(t) = R_y(0, t) = r_1 \sum_{m=-\infty}^{\infty} [(1 - r_1)g^2(t - mT)$$
$$+ (1 - 2r_1)g(t - mT)g[t - (m + 1)T]. \tag{77}$$

For the independent binary message ensemble with $g(t) = g_0(t)$, $0 < t < T$,

$$\text{av } y^2(t) = R_y(0, t) = m_1(1 - m_1) \sum_{m=-\infty}^{\infty} g_0^2(t - mT)$$
$$= \frac{m_1(1 - m_1) \exp [-\omega_0(t - T/2)/Q]}{4 \sinh \omega_0 T/2Q (\cosh^2 \omega_0 T/2Q - \cos^2 \omega_0 T)} \tag{78}$$
$$\cdot \{\cosh^2 \omega_0 T/2Q - \cos^2 \omega_0 T + \sinh^2 (\omega_0 T/2Q) \cos \omega_0 T$$
$$\cdot \cos [\omega_0(2t - T) + 2\theta] - \sinh (\omega_0 T/Q) \cosh (\omega_0 T/2Q)$$
$$\cdot \sin \omega_0 T \sin [\omega_0(2t - T) + 2\theta]\}.$$

If $\omega_0 T = 2\pi$, this reduces to:

$$\text{av } y^2(t) =$$
$$\frac{m_1(1 - m_1)\omega_r^2(4Q^2 + 1) \exp [-\omega_r(t - T/2)/Q] \cos^2 (\omega_r t + \theta)}{8R^2Q^4 \sinh \pi/Q}. \tag{79}$$

In the limit as Q approaches infinity, then,

$$\text{av } y^2(t) \rightarrow \frac{m_1(1 - m_1)\omega_r^2}{2\pi R^2 Q} \cos^2 \omega_r t. \tag{80}$$

IV. EVALUATION OF RMS PHASE ERROR WHEN INDEPENDENT ON-OR-OFF UNIT IMPULSES ARE APPLIED TO MISTUNED CIRCUIT

We illustrate the use of the equations derived above by carrying through the evaluation of (62) when the message pulse source consists of independent on-or-off unit impulses and the ratio of tuning error to the

signaling rate is small compared with unity. We adopt the following notation:

$$k(\omega_0 - \omega_r)/\omega_r = \text{fractional tuning error,}$$
$$\alpha = \omega_0 T/4Q = (1 + k)\pi/2Q,$$
$$\beta = (1 + k)\pi,$$

(81)

$$\omega_0[(t_c - T/2) + \theta] = \Phi.$$

From (69), then,

$$\sin \Phi = \tanh \alpha \cot \beta \, (1 + \tanh^2 \alpha \cot^2 \beta)^{-1/2},$$
$$\cos \Phi = (1 + \tanh^2 \alpha \cot^2 \beta)^{-1/2}.$$

(82)

Then, from (78), we may write, after noting that av $x(t_c) = 0$ implies av $x^2(t_c) = $ av $y^2(t_c)$:

av $x^2(t_c) =$

$$\frac{m_1(1 - m_1)\omega_0^2(4Q^2 + 1) \exp[-\omega_0(t_c - T/2)/Q]}{16R^2Q^4 \sinh 2\alpha\,(\cosh^2 2\alpha - \cos^2 2\beta)} \{\sinh^2 2\alpha$$

(83)

$$+ \sin^2 2\beta + \sinh^2 2\alpha \cos 2\beta \cos 2\Phi - \sinh 2\alpha \cosh 2\alpha \sin 2\beta \sin 2\Phi\}.$$

From straightforward trigonometrical manipulations,

$$\sin 2\Phi = \frac{2 \tanh \alpha \cot \beta}{1 + \tanh^2 \alpha \cot^2 \beta},$$
$$\cos 2\Phi = \frac{1 - \tanh^2 \alpha \cot^2 \beta}{1 + \tanh^2 \alpha \cot^2 \beta}.$$

(84)

When the values of $\sin 2\Phi$ and $\cos 2\Phi$ are substituted in (83) we find, after some reduction:

av $x^2(t_c) =$

$$\frac{m_1(1 - m_1)\omega_0^2(4Q^2 + 1) \sin 2\beta \qquad \exp\{-Q^{-1}[\theta + \text{arc tan}\,(\tanh \alpha \cot \beta)]\}}{16R^2Q^4 \sinh 2\alpha\,(\cosh^2 2\alpha - \cos^2 2\beta)}.$$

(85)

In the high-Q case, the value of av $x(t_c + T/4)$ will be very nearly equal to the maximum value of av $x(t)$ as given by (68). Making this assumption and simplifying, we obtain:

av^2 $x(t_c + T/4) \doteq$

$$\frac{m_1^2\omega_0^2(4Q^2 + 1) \exp\{-Q^{-1}[\omega_0/4f_r + \theta + \text{arc tan}\,(\tanh \alpha \cot \beta)]\}}{16R^2Q^2(\cosh^2 \alpha - \cos^2 \beta)}.$$

(86)

Then, from (62),

$$\text{av } \epsilon^2 \doteq \frac{(1 - m_1) \sin^2 2\beta \exp(-\omega_0/4 f_r Q)}{4m_1 \sinh 2\alpha (\cosh^2 \alpha - \sin^2 \beta)}$$

$$= \frac{(1 - m_1) \sin^2 2k\pi \exp[-(1 + k)\pi/2Q]}{4m_1 \sinh(1 + k)\pi/Q [\cosh^2(1 + k)\pi/2Q - \sin^2 k\pi]}. \tag{87}$$

In the limit for $k \ll 1$ and $Q \gg 1$, which is the region of practical interest,

$$\text{av } \epsilon^2 \to (1 - m_1)\pi k^2 Q/m_1. \tag{88}$$

The rms phase error is the square root of this quantity. For equal probability of pulses and spaces, $m_1 = \frac{1}{2}$, and

$$\epsilon_{rms} = (\text{av } \epsilon^2)^{1/2} \doteq k\sqrt{\pi Q}. \tag{89}$$

Fig. 9 shows a set of curves of rms phase error vs. Q for fixed values of tuning error. It is to be noted that this performance is attained when the reference phase of negative axis-crossings is t_c defined by av $x(t_c) = 0$. This value of time is displaced from the time t_0 defined by (42), which would be the correct reference if there were no mistuning. The difference is accounted for by the steady-state phase shift of the network at the
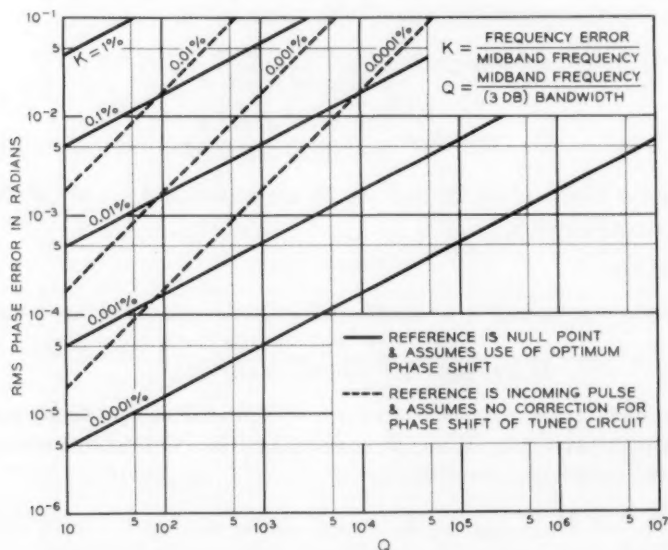

Fig. 9 — Phase error in timing wave from mistuned recovery circuit.

pulse frequency. If we were to take the axis-crossing reference time as t_0, the mean square phase error would be determined by av $x^2(t_0)$ instead of av $x^2(t_c)$ in (62). Since av $x(t_0)$ does not vanish, the expression for av $x^2(t_0)$ must be written as

$$\text{av } x^2(t_0) = \text{av}^2 \, x(t_0) + \text{av } y^2(t_0). \tag{90}$$

From (42),

$$t_0 = T/4 - T\theta/2\pi,$$

$$t_0 - T/2 = -T/4 - \theta/\omega_r = -(\theta + \pi/2)/\omega_r,$$

$$\cos[\omega_0(t_0 - T/2) + \theta] = \cos[(1+k)\pi/2 - k\theta]$$

$$= -\sin[k(\pi/2 - \theta)],$$

$$\sin[\omega_0(t - T/2) + \theta] = \cos[k(\pi/2 - \theta)], \tag{91}$$

$$\cos 2[\omega_0(t_0 - T/2) + \theta] = -\cos k(\pi - 2\theta),$$

$$\sin 2[\omega_0(t - T/2) + \theta] = -\sin k(\pi - 2\theta).$$

Substituting these values and taking the limit for $k \ll 1$, and $Q \gg 1$, we find the rms phase error in the region of interest to be given approximately by

$$\epsilon'_{rms} \doteq \frac{k(1 + 1/4Q)}{(k^2 + 1/4Q^2)^{1/2}} \doteq 2Qk. \tag{92}$$

The dashed curves on Fig. 9 show the values of error thus obtained and indicate the importance of a phase adjustment to optimize for the tuned circuit in use.

It remains to evaluate the variation in peak height of the tuned circuit response. This requires evaluating av ζ^2 from (64), which in turn requires calculation of av $y^2(t_c + T/4)$ from (78). As in the case of av^2 $y(t_c + T/4)$ it is sufficient to take the maximum value of (78) over a period in t. The evaluation of av ζ^2 is facilitated by immediate comparison of the resulting expression for av^2 $y(t_c + T/4)$ with that for av^2 $x(t_c + T/4)$ given by (86). We find

$$\text{av } y^2(t_c + T/4) = \text{av}^2 \, x(t_c + T/4) \frac{B(1 - m_1)(\cosh^6 \alpha - \cosh^2 \beta)}{m_1 \sinh^2 \alpha \, (\cosh^6 2\alpha - \cos^2 2\beta)}, \tag{93}$$

where

$$B = \cosh^2 2\alpha - \cos^2 2\beta$$

$$+ \sinh 2\alpha[\sinh^2 2\alpha \cos^2 2\beta + \cosh^2 2\alpha \sin^2 2\beta]^{1/2}. \tag{94}$$

Then, by some further manipulation, we find:

$$\text{av } \zeta^2 =$$

$$\frac{(1 - m_1)[\sinh^2 2\alpha + \sin^2 2\beta + \sinh 2\alpha (\sinh^2 2\alpha + \sin^2 2\beta)^{1/2}]}{4m_1 \sinh 2\alpha (\cosh^2 \alpha - \sin^2 \beta)}. \quad (95)$$

In the limit when $Q \gg 1$ and $k \ll 1$, we obtain

$$\text{av } \zeta^2 = \frac{(1 - m_1)\pi Q}{4m_1(1 - k^2\pi^2)} \left[\frac{1}{Q^2} + 4k^2 + \frac{1}{Q}\left(\frac{1}{Q^2} + 4k^2\right)^{1/2} \right]. \quad (96)$$

With no tuning error, the ratio

$$\text{av } \zeta^2 \rightarrow \frac{(1 - m_1)\pi}{2m_1 Q} \quad (97)$$

can also be obtained by taking the ratio of (80) to the square of (71). When $k \ll \frac{1}{2}Q$, the limiting form is

$$\text{av } \zeta^2 \doteq \frac{(1 - m_1)\pi}{2m_1 Q} (1 + 3k^2 Q^2). \quad (98)$$

If k is fixed and not equal to zero, the limit as Q is made very large is

$$\text{av } \zeta^2 = \frac{(1 - m_1)\pi Q k^2}{m_1}. \quad (99)$$

The analysis given here is valid only when $\text{av } \zeta^2 \ll 1$.

V. EFFECT OF NOISE ON TIMING

The question of timing errors caused by noise superimposed on the received pulse train can be resolved into two parts: (1) the shift in reference time at the input to the tuned circuit and (2) the resultant variation
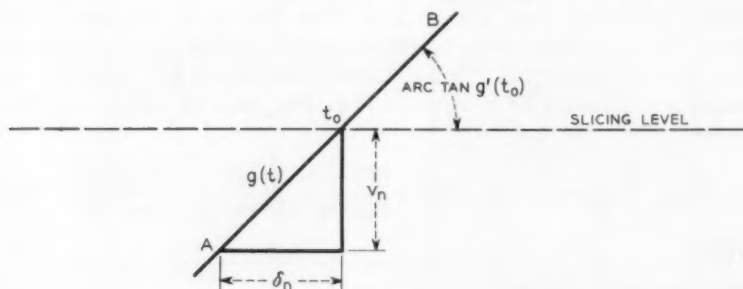

Fig. 10 — Geometrical construction showing relation between noise and timing error.

in the phase of the tuned circuit response. The first question can be conveniently answered by a geometrical argument based on Fig. 10. Here an enlarged section of the message pulse wave front is approximated by a straight line AB of slope equal to that of the pulse at the unperturbed critical time t_0. The instantaneous noise voltage v_n advances the critical time by δ_n, where v_n/δ_n = slope, and hence

$$\delta_n = v_n/g'(t_0). \tag{100}$$

Statistics of δ_n can be evaluated from the statistics of the noise and the specification of the pulse waveform. In particular, if the instantaneous noise values at corresponding instants of signaling intervals n intervals apart are not independent but have an autocorrelation or autocovariance function $R_v(n)$, the resulting time shifts have the corresponding function $R_\delta(n)$, where

$$R_\delta(n) = \text{av} \ (\delta_k\delta_{k+n}) = \text{av} \ (v_kv_{k+n})/g'(t_0). \tag{101}$$

The average is taken over all time for one member of the ensemble to give the autocorrelation or over the ensemble at fixed time to give the autocovariance. We assume here that the process is ergodic and hence that the two averages are equal. We have also assumed that the slope of the signal pulse is constant in the neighborhood of the slicing level.

In the second part of the problem, we assume perfect tuning of the recovery circuit and write for its response

$$z(t) = \sum_{n=-\infty}^{\infty} a_ng_0(t - nT + \delta_n). \tag{102}$$

Here $g_0(t)$ represents any response function which reproduces the axis crossings correctly and, in particular cases evaluated, it will be assumed as the impulse response of a tuned circuit. We assume that the individual values of δ_n are sufficiently small to enable an accurate representation by first-order terms in a power series, thus:

$$z(t) \doteq \sum_{n=-\infty}^{\infty} a_n[g_0(t - nT) + \delta_ng_0'(t - nT)]$$

$$= x(t) + \sum_{n=-\infty}^{\infty} a_n\delta_ng_0'(t - nT). \tag{103}$$

Without loss of generality, we may take av $\delta_n = 0$, since a constant time shift may be compensated by a change in the reference time. Then, if the noise and message ensembles are independent,

$$\text{av} \ z(t) = \text{av} \ x(t). \tag{104}$$

We next calculate the autocovariance of the perturbed tuned circuit response as

$$\text{av } [z(t)z(t + \tau)] = \text{av } [x(t)x(t + \tau)]$$
$$+ \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \text{av } (a_m a_n) \text{ av } (\delta_m \delta_n) g_0'(t - mT) g_0'(t - nT + \tau). \quad (105)$$

The first term on the right represents the unperturbed autocovariance and has already been evaluated. The effect of the noise is given by the second term and, by use of the same technique as in obtaining (16), we obtain its autocovariance as

$$R_{z-x}(\tau, t) =$$
$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} R(m - n) R_\delta(m - n) g_0'(t - mT) g_0'(t - nT + \tau). \quad (106)$$

In particular for the case of independent message values,

$$R_{z-x}(\tau, t) = m_1 \sum_{m=-\infty}^{\infty} R_\delta(m) g_0'(t - mT) g_0'(t - mT + \tau), \quad (107)$$

and

$$\text{av } [z^2(t) - x^2(t)] = R_{z-x}(0, t) = m_1 \sum_{m=-\infty}^{\infty} R_\delta(m) [g_0'(t - mT)]^2. \quad (108)$$

We may treat the phase error caused by noise in the same way as we did the mistuning error in deriving (62). Since, with no tuning error, av $x^2(t_c) = 0$ and $t_c = T/4 - \theta/2\pi f_r$,

$$\text{av } \epsilon_n^2 = \frac{\text{av } z^2(t_c)}{\text{av}^2 x(t_c + T/4)}. \quad (109)$$

If the input noise values are assumed independent,

$$R_\delta(m) = \text{av } \delta_n^2 = \text{av } v_n^2/[g'(t_0)]^2 \equiv \delta^2. \quad (110)$$

Then

$$\text{av } z^2(t_c) = m_1 \delta^2 \sum_{m=-\infty}^{\infty} [g_0'(t_c - mT)]^2. \quad (111)$$

If

$$g_0(t) = Ae^{-\alpha t} \cos(\omega_r t + \theta), \quad (112)$$

it follows that

$$g_0'(t) = Be^{-\alpha t} \cos(\omega_r t + \theta - \Phi), \quad (113)$$

where

$$B = (\alpha^2 + \omega_r^2)^{1/2} A, \qquad \tan \Phi = \omega_r/\alpha. \qquad (114)$$

Hence, the previous evaluation of (79) can be used for av $z^2(t_c)$, replacing θ by $\theta - \Phi$ and multiplying the result by $(\alpha^2 + \omega_r^2)\delta^2$. Then

$$\text{av } z^2(t_c) = \frac{m_1\omega_r^4\delta^2(4Q^2 + 1) \exp (\theta + \pi)/2Q}{R^2Q^4 \sinh \pi/Q}. \qquad (115)$$

The limit as Q approaches infinity is

$$\text{av } z^2(t_c) = \frac{m_1\omega_r^4\delta^2}{2\pi R^2Q}. \qquad (116)$$

The expression for av ϵ_n^2 is found from division by the square of the amplitude factor of (70), giving

$$\text{av } \epsilon_n^2 = m_1^{-1} \omega_r^2\delta^2 \exp [(3\pi - 2\theta)/4Q] \tanh (\pi/2Q)$$

$$\rightarrow \frac{\omega_r^2\pi}{2m_1Q}\delta^2 \quad \text{as} \quad Q \rightarrow \infty \qquad (117)$$

The effect of noise on the recovered timing is seen to vary inversely with Q and hence may be made arbitrarily small by using a sufficiently high-Q circuit. The effect is opposite to that of tuning error which becomes worse as Q is increased. The possibility of an optimum Q taking both effects into account thus exists.

VI. EFFECT OF TIMING VARIATIONS ON A DECODED ANALOG SIGNAL

The effects of timing errors are of two main kinds. First, they increase the difficulty of reading the message correctly because the decisions are made at a less favorable point on the pulse wave. This is called the alignment error and is particularly important in a long chain of regenerative repeaters. It has been discussed by E. D. Sunde[4] and is also analyzed in the companion paper by H. E. Rowe.[1] It will not be treated here. The second effect is to impair the usefulness of the received digital message even though the correct sequence of values is delivered. The seriousness of this effect varies widely with the type of message. At one extreme, the recovery of printed text from a telegraph message would hardly be affected. An analog signal transmitted by pulse code modulation, on the other hand, is in some danger because the decoded signal is not correctly reproduced by irregularly spaced samples. A single-channel system would undergo phase modulation from this cause. In the case of a time-division multiplex system, some interchannel crosstalk may result if the jitter varies with the signals in the channels. The most severe requirement oc-

curs when PCM is applied to transmit a group of frequency-multiplexed channels, since any waveform distortion can mean interchannel crosstalk after the channels are separated.

We should point out at the beginning that we are not forced to accept time jitter in the received digital wave as finally used. A master clock could be inserted at any stage to force the pulse back into a proper time-reference framework. By means of resampling combined with pulse stretching and delay line distribution techniques, such corrections can be made even when the extent of the tuning variations exceeds the signaling interval. It would, of course, be desirable to avoid these operations and hence it is important to determine the permissible extent of time jitter in specific situations. We shall consider here the case of multiplex speech channels in frequency division transmitted by pulse code modulation, the FDM-PCM system.

Our procedure will be to represent the multiplex signal by an ensemble of band-limited functions $\{s(t)\}$ having a spectral density $w_s(f)$ which vanishes for $|f| \geq f_0$. Any member of the ensemble is completely defined by its samples taken $1/2f_0$ apart; in fact, by the sampling theorem, we have the identity:

$$s(t) = \sum_{n=-\infty}^{\infty} s(n/2f_0) \frac{\sin 2\pi f_0(t + n/2f_0)}{2\pi f_0(t + n/2f_0)}. \tag{118}$$

The right-hand member is the response of an ideal low-pass filter to impulses of weight proportional to the samples. The effect of time jitter is to introduce an irregular spacing of the recovered samples which, when applied to an ideal filter, produce the distorted wave

$$\sigma(t) = \sum_{n=-\infty}^{\infty} s(n/2f_0) \frac{\sin 2\pi f_0(t + u_n + n/2f_0)}{2\pi f_0(t + u_n + n/2f_0)}. \tag{119}$$

The timing errors are represented by $u_n = u(n/2f_0)$, where $\{u(t)\}$ represents a jitter ensemble with specified statistical properties.

The evaluation of interchannel crosstalk is conveniently accomplished by assuming signals present in all but one of the FDM channels. The spectral density $w_s(f)$ is set equal to zero throughout the frequencies occupied by the idle channel. Values in this frequency range of $w_\sigma(f)$, the spectral density of the recovered signals, then give the interchannel interference caused by the time jitter. This method has the advantage of separating the interchannel interference sought from in-band distortion which is of no consequence. We refer to a previous paper[6] for a description of some of the necessary mathematical techniques.

We assume that $s(t)$ and $u(t)$ have autocovariance functions expressed in terms of the spectral densities by:

$$R_s(\tau) = \int_{-f_0}^{f_0} w_s(f) \exp\ (j2\pi\tau f)\ df, \tag{120}$$

$$R_u(\tau) = \int_{-f_0}^{f_0} w_u(f) \exp\ (j2\pi\tau f)\ df. \tag{121}$$

It is no loss of generality to assume that $u(t)$ is band-limited to the range $-f_0$ to f_0, since higher frequency components would produce effects indistinguishable from those in the band of half the sampling rate. The spectral densities are in turn expressible in terms of the autocovariances by

$$w_s(f) = \int_{-\infty}^{\infty} R_s(\tau) \exp\ (-j2\pi f\tau)\ d\tau, \tag{122}$$

$$w_u(f) = \int_{-\infty}^{\infty} R_u(\tau) \exp\ (-j2\pi f\tau)\ d\tau. \tag{123}$$

The s- and u- ensembles are assumed to be independent and ergodic. Note also that the signal density spectrum is defined in terms of mean squared voltage or current values, while the jitter spectrum is expressed in terms of mean square values of time.

Our procedure is to calculate first the autocorrelation function of the $\sigma(t)$- ensemble by the definition

$$R_\sigma(\tau) = \text{av}\ \sigma(t)\sigma(t + \tau) = \text{av} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} s(m/2f_0)s(n/2f_0)$$
$$\cdot\ \frac{\sin 2\pi f_0(t + u_m + m/2f_0) \sin 2\pi f_0(t + \tau + u_n + n/2f_0)}{4\pi^2 f_0^2(t + u_m + m/2f_0)(t + \tau + u_n + n/2f_0)}\ . \tag{124}$$

The spectral density of the distorted signal ensemble is then given by

$$w_\sigma(f) = \int_{-\infty}^{\infty} R_\sigma(\tau) \exp\ (-j2\pi f\tau)\ d\tau. \tag{125}$$

We assume ensemble and time averages to be interchangeable by the ergodic assumption. The averaging over the s- and u- ensembles may be done separately because of their independence. Furthermore, the double series may be averaged term by term since, in any case, the average of the sum is equal to the sum of the individual averages. We first note that, for the s- ensemble averaging,

$$\text{av}\ [s(m/2f_0)s(n/2f_0)] = \text{av}\ [s(m/2f_u)s(m + n - m)/2f_0]$$
$$= R_s[(n - m)/2f_0] = R_s[(m - n)/2f_0]. \tag{126}$$

The averaging over the u- ensemble is more troublesome.

It has been found possible to obtain a reasonably simple answer only in the case of a gaussian jitter ensemble. By an involved calculation, which is explained in the Appendix, we obtain for this case, setting $T_0 = 1/2f_0$,

$$R_\sigma(\tau) = \frac{T_0}{\pi} \sum_{n=-\infty}^{\infty} R_s(nT_0) \int_0^{\pi/T_0} \exp\left(-z^2 U_n^2\right) \cos\left(\tau - nT_0\right)z \, dz, \quad (127)$$

where

$$U_n^2 = R_u(0) - R_u(nT_0). \quad (128)$$

In terms of the complex error function,

$$R_\sigma(\tau) = \frac{T_0}{2\sqrt{\pi}} \sum_{n=-\infty}^{\infty} \frac{R_s(nT_0)}{U_n} \operatorname{Re} \operatorname{erf}\left(\frac{\pi U_n}{T_0} - i\frac{\tau + nT_0}{2U_n}\right). \quad (129)$$

In the case of PCM transmission of an FDM group, the allowable jitter is small. This means that the integral of the spectral density $w_u(f)$ must be small and, hence, that $R_u(0)$, which is equal to this integral, must be small; $R_u(\tau)$ assumes its largest values at $\tau = 0$ and we conclude *a fortiori* that U_n^2 must be small. Hence, we approximate by replacing $\exp\left(-z^2 U_n^2\right)$ by zero and first-power terms in its power series expansion and obtain an approximate value of the integral in (127):

$$R_\sigma(\tau) \doteq \sum_{n=-\infty}^{\infty} R_s(nT_0) \frac{\sin 2\pi f_0(\tau + nT_0)}{2\pi f_0(\tau + nT_0)} (t - 4\pi^2 f_0^2 U_n^2)$$
$$- 2U_n^2 \left[\frac{\cos 2\pi f_0(\tau + nT_0)}{(\tau + nT_0)^2} - \frac{T_0 \sin 2\pi f_0(\tau + nT_0)}{\pi(\tau + nT_0)}\right]. \quad (130)$$

By a further manipulation also given in the Appendix, we find that this approximation yields for the spectral density

$$w_\sigma(f) \doteq (1 - 4\pi^2 f^2 u_0^2) w_s(f)$$
$$+ 4\pi^2 f^2 [w_{s*u}(f) + w_{s*u}(2f_0 + f) + w_{s*u}(2f_0 - f)]; \; |f| < f_0, \quad (131)$$

where

$$u_0 = \sqrt{R_y(0)} \text{ is the rms time jitter and} \quad (132)$$

$$w_{s*y}(f) = \int_{-\infty}^{\infty} w_s(\lambda + f) w_u(\lambda) \, d\lambda \quad (133)$$

$$= \text{convolution of signal and jitter spectrum.}$$

We note that the first term consists of the original power spectrum of the signal depressed by the "crowding" effect factor $(2\pi f u_0)^2$. The crowd-

ing effect does not introduce new frequencies not present in the original signal and therefore does not contribute to interchannel interference. The interference is given entirely by the second term, which consists of differentiated convolution spectra of signal and jitter. The differentiation is expressed by the multiplying factor $(2\pi f)^2$, which gives a triangular voltage spectrum like that of thermal noise in frequency modulation. The term $w_{s \cdot y}(f)$ represents the direct convolution of signal and jitter spectra, while $w_{s \cdot u}(2f_0 \pm f)$ represents sidebands of the convolution spectrum on the sampling frequency.

In testing of multichannel systems by noise-band loading, as previously mentioned, the signal spectrum applied is flat except for the narrow band assigned to the channel under test. The power spectrum is made equal to zero throughout the test-channel frequencies (see Figs. 11 and 12). The expression for $w_s(f)$ is then:

$$w_s(f) = \begin{cases} K; & |f| < f_t \quad \text{and} \quad f_t + f_c < |f| < f_0 \\ 0; & f_t < |f| < f_t + f_c, \end{cases} \tag{134}$$

where f_t is the test channel carrier frequency and f_c is the bandwidth of one channel. (We have assumed a single sideband system with upper sideband transmitted.) Then (133) gives

$$w_{s \cdot u}(f) = K \left[\int_{-f_0}^{-f_t - f_c} + \int_{-f_t}^{f_t} + \int_{f_t + f_c}^{f_0} \right] w_u(f + \lambda) \, d\lambda. \tag{135}$$

It is sufficient to evaluate the interference spectrum in the test channel range f_t to $f_t + f_c$ and $-f_t - f_c$ to $-f_t$. Since power spectra are even
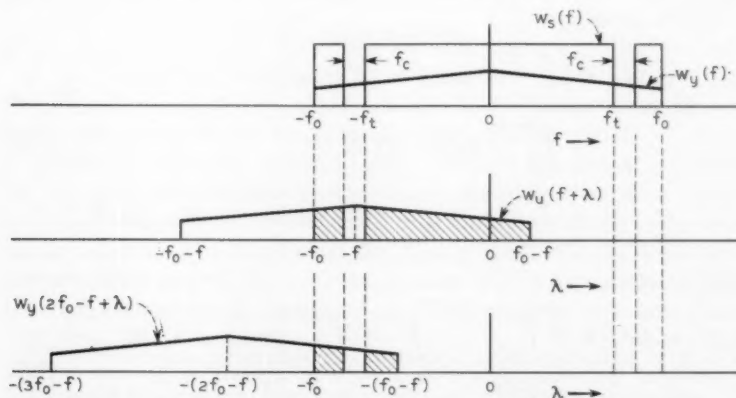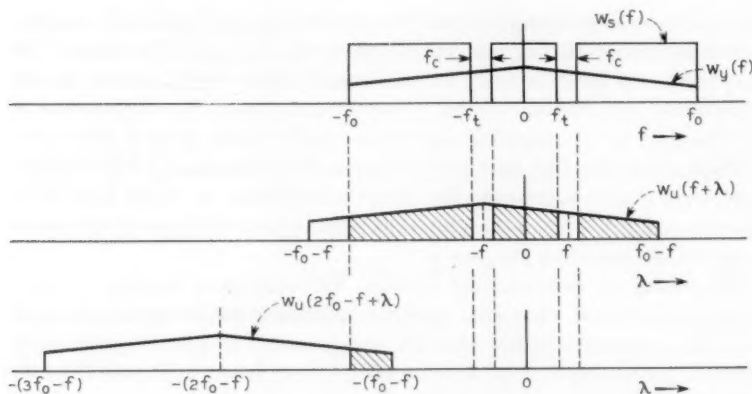

Fig. 11 — Spectral relations for noise band as test signal with $f_t > f_0/2$.

Fig. 12 — Spectral relations for noise band as test signal with $f_t < f_0/2$.

functions, we actually need to calculate for only one of these intervals. There are two contributions from (131) to the positive interval — one for $w_{s \cdot u}(f)$ and the other from $w_{s \cdot u}(2f_0 - f)$. The term $w_{s \cdot u}(2f_0 + f)$ vanishes for $f > 0$ since $w_{s \cdot u}(\nu)$ is zero for $\nu > 2f_0$.

The contribution from $w_{s \cdot u}(2f_0 - f)$ to the positive interval f_t to $f_t + f_c$ is defined by the interval $2f_0 - f = f_t + f_c$ to $2f_0 - f = f_t$ or $f = 2f_0 - f_t - f_c$ to $f = 2f_0 - f_t$. We must therefore evaluate $w_{s \cdot u}(f)$ in the ranges f_t to $f_t + f_c$ and $2f_0 - f_t - f_c$ to $2f_0 - f_t$.

Figs. 11 and 12 show the limits of integration which apply for the cases of f_t greater than and less than $f_0/2$, respectively. From these diagrams, we deduce that the distortion spectrum in the range $f_t < f < f_t + f_c$, is given by

$$
\begin{aligned}
w_{sd}(f) = 4\pi^2 f^2 K \bigg\{ &\int_{-f_0}^{-f_t - f_c} [w_u(f + \lambda) + w_u(2f_0 - f + \lambda)] \, d\lambda \\
&+ \int_{-f}^{f_0 - f_t} w_u(f + \lambda) \, d\lambda \\
&+ \int_{-f_t}^{f - f_0} w_u(2f_0 - f + \lambda) \, d\lambda \bigg\} \qquad \text{for } f > \frac{f_0}{2},
\end{aligned}
\tag{136}
$$

and

$$
\begin{aligned}
w_{sd}(f) = 4\pi^2 f^2 K \bigg\{ &\bigg[\int_{-f_0}^{-f_t - f_c} + \int_{-f_t}^{f_t} + \int_{f_t + f_0}^{f_0 - f} \bigg] w_u(f + \lambda) \, d\lambda \\
&+ \int_{-f_0}^{f - f_0} w_u(2f_0 - f + \lambda) \, d\lambda \bigg\} \qquad \text{for } f < \frac{f_0}{2}.
\end{aligned}
\tag{137}
$$

A special case of interest is that of a flat jitter band over a low-frequency portion of the signal band. Rowe's previously cited work[1] on regenerative repeater chains indicates that the jitter spectrum peaks at the low-frequency edge of the signal band as the number of repeaters becomes large. Suppose the jitter spectrum is uniform from $-F$ to F, $f < f_0$ and consider the case of $f_t > f_0/2$ and $F < f_0 - f_t$. This corresponds to a test channel in the upper half of the signal band but below the top by at least the width of the jitter band. The distortion spectrum is given by

$$
\begin{aligned}
w_{\sigma d}(f) &= \frac{4\pi^2 f^2 K u_0^2}{2F} \left(\int_{-f-F}^{-f_t-f_c} + \int_{-f_t}^{-f+F} \right) d\lambda \\
&= \frac{4\pi^2 f^2 K u_0^2}{2F} \left(-f_t - f_c + f + F - f + F + f_t \right) \qquad (138) \\
&= \frac{4\pi^2 f^2 K u_0^2}{2F} \left(2F - f_c \right) \\
&\doteq 4\pi^2 f^2 K u_0^2 \qquad\qquad\qquad \text{if } f_c \ll 2F.^*
\end{aligned}
$$

The mean square total interference voltage in the idle channel is

$$
u_c^2 = 2 \int_{f_t}^{f_t+f_c} w_{\sigma d}(f)\, df \doteq 8\pi^2 f_t^2 K u_0^2 f_c. \qquad (139)
$$

If the mean square signal is s_0^2, we have

$$
s_0^2 = 2K f_0, \qquad (140)
$$

and

$$
\frac{u_c^2}{s_0^2} = 4\pi^2 f_t^2 u_0^2 f_c / f_0. \qquad (141)
$$

The dependence on the square of test-channel frequency shows that we have a triangular voltage spectrum like that obtained in FM.

A quantity of interest defining signal quality in the typical channel of the frequency-division multiplex system is the ratio M of mean square value of the sine wave test tone which fully loads the channel to the mean square interference voltage when the channel is idle and the other channels are being operated in a normal manner. The mean square full load test tone voltage on one channel of an FDM telephone system is related

---

* The implications of this approximation should not be overlooked. If the bandwidth of the jitter spectrum is small compared to a channel width, interchannel interference would result only from high-order modulation neglected in the above approximation.

to the mean square full load test tone for the entire system by the Hol-brook-Dixon[7] curves, which we shall express in the form

$$M_n = n\, P_1/P_n ,\tag{142}$$

where

P_1 = full load test tone mean square value necessary to transmit one
      voice channel,
P_n = full load test tone mean square value necessary to transmit n
      superimposed voice channels.

When the multiplex system is operating normally the active channels are loaded with voice signals and not by sine waves. The value of s_0^2, the mean square total signal voltage with speech loading, may be expressed as

$$s_0^2 = P_n/H^2 ,\tag{143}$$

where H is the ratio of the peak factor of a sine wave to that of a composite multichannel speech wave. This is based on the assumption that the normal speech loading reaches system overload occasionally. The peak factor of a sine wave is $\sqrt{2}$, corresponding to 3 db, while the peak factor of superimposed speech channels approaches that of thermal noise as the number of channels is made large. The peak factor of thermal noise depends on the value of the probability of excess chosen. A value often used is 4, corresponding to 12 db, for which the probability of observing greater peaks is one in 10,000. This choice fixes H^2 at eight, corresponding to 9 db.

We write, in accordance with the above,

$$M = P_1/u_c^2 ,\tag{144}$$

and assume M specified. It follows that

$$M = \frac{M_n P_n f_0}{4\pi^2 f_i^2 u_0^2 n f_c s_0^2} = \frac{H^2 M_n f_0}{4\pi^2 f_i^2 u_0^2 n f_c} .\tag{145}$$

Solving for u_0^2, we obtain the minimum requirement for mean square time jitter:

$$u_0^2 = \frac{H^2 f_0 M_n}{4\pi^2 f_i^2 f_c n M} .\tag{146}$$

Because of the triangular distortion voltage spectrum, the requirement is most severe at the highest channel carrier frequency.

It is convenient to express this requirement in terms of rms phase

jitter Φ allowed at the digital pulse frequency, which is $2Nf_0$ for an N-digit PCM system. Since $\Phi = 2\pi(2Nf_0)u_0$,

$$\Phi^2 = \frac{4N^2H^2f_0^3M_n}{nf_t^2f_cM}. \tag{147}$$

Consider, for example, an 8-digit binary PCM system transmitting a 2000-channel single-sideband FDM telephone signal with 4-kc carrier spacing. The top channel is at 8 mc and imposes the most severe requirement. We ask for a 60-db ratio of full load test tone power to idle channel interference. Assume the sampling frequency is 20 mc. We evaluate M_n by extrapolating from Table I of Ref. 8: $H^2 = 8$, $N = 8$, $f_0 = 10^7$, $M_n = 2000/80$, $n = 2000$, $f_t = 8 \times 10^6$, $f_c = 4 \times 10^3$, $M = 10^6$, and we calculate $\Phi = 0.3$ radian. This result is close to estimates made by O. E. DeLange and E. D. Sunde by different methods.

VII. ACKNOWLEDGMENT

The work described here had its inception in the work on PCM done at Bell Laboratories in the early 1940's and its beginnings were influenced by discussions with C. B. Feldman, R. L. Dietzold and L. A. MacColl. Later benefits were received by comparison with results obtained on various features of the problem by J. R. Pierce, H. E. Rowe and E. D. Sunde. The present form has profited from suggestions by D. W. Tufts.

APPENDIX

By the definition of a time average,

$$R_\sigma(\tau) = \lim_{N\to\infty} \frac{1}{(2N+1)T_0} \sum_{m=-N}^{N} \sum_{n=-N}^{N} R_z[(m-n)T_0]$$

$$\int_{-NT_0}^{NT_0} \frac{\sin\frac{\pi}{T_0}(t+mT_0+u_m)\sin\frac{\pi}{T_0}(t+\tau+nT_0+u_n)\,dt}{\left(\frac{\pi}{T_0}\right)^2(t+mT_0+u_m)(t+\tau+nT_0+u_n)}. \tag{148}$$

The integrand is an analytic function of t in the finite plane and hence the path of integration may be deformed without changing the value of the integral. We replace the part of the path passing through the zeros of the denominator by downward indentations. Sum and difference formulas may then be applied to the sines to resolve the integral into the sum of separate components without introducing singularities in the inte-

grands. Designating the resulting path of integration by C, we then have:

$$R_x(\tau) = \lim_{N=\infty} \sum_{m=-\infty}^{N} \sum_{n=-\infty}^{N} \frac{T_0 R_z[(m-n)T_0]}{4N\pi^2}$$

$$\int_c \frac{dt}{(t + mT_0 + u_m)(t + \tau + nT_0 + u_n)}$$

$$\left\{ \cos \frac{\pi}{T_0} [\tau + (n-m)T_0 + u_n - u_m] \right.$$

$$\left. - \cos \frac{\pi}{T_0} [2t + \tau + (m+n)T + u_m + u_n] \right\}. \tag{149}$$

To evaluate the integral, write:

$$\left. \begin{array}{r} mT_0 + u_m = a \\ \tau + nT_0 + u_n = b \end{array} \right\}. \tag{150}$$

Note that a and b are real numbers. We must calculate:

$$\int_c \frac{\cos \frac{\pi}{T_0}(b-a)}{(t+a)(t+b)} dt - \int_c \frac{\cos \frac{\pi}{T_0}(2t+a+b)}{(t+a)(t+b)} dt.$$

The first integral vanishes as we let N approach infinity, as may be seen by closing the contour in an infinite semicircle below the real t-axis. The absolute value of the integral around the semicircle of radius r cannot exceed the product of length of path πr and the maximum absolute value of the integrand, which is less than $1/r^2$. The integral around the semicircle is therefore less in absolute value than π/r and approaches zero as r goes to infinity. The integral around the closed contour including C and the semicircle must vanish because there are no singularities inside. Hence the integral over C also vanishes and the first integral is zero.

The second integral may be written as the sum of two terms:

$$\frac{e^{j\pi(a+b)/T_0}}{2} \int_c \frac{e^{j2\pi t/T_0} dt}{(t+a)(t+d)} + \frac{e^{-j\pi(a+b)/T_0}}{2} \int_c \frac{e^{-j2\pi t/T_0} dt}{(t+a)(t+b)}. \tag{151}$$

In the first term we close the contour in an infinite semicircle above the real t-axis and obtain an integrand which vanishes exponentially as the radius is made large. In the limit the integral along C is then equal to the integral around the closed contour, which in turn is equal to $2\pi/j$ times the sum of the residues at the poles $t = -a$ and $t = -b$. In the second

term we close the contour below the real t-axis and show that the integral vanishes. Therefore the complete result is given by:

$$-\frac{2\pi j}{2} e^{j\pi(a+b)/T_0} \left(\frac{e^{-j2\pi a/T_0}}{b-a} + \frac{e^{-j2\pi b/T_0}}{a-b} \right)$$

$$= -\frac{\pi j}{a-b} [e^{j\pi(a-b)T_0} - e^{-j\pi(a-b)/T_0}] \quad (152)$$

$$= \frac{2\pi \sin \pi(a-b)/T_0}{a-b}.$$

Hence,

$$R_\sigma(\tau) =$$

$$\lim_{N\to\infty} \sum_{n=-N}^{N} \sum_{b=-N}^{N} \frac{T_0 R_s[(m-n)T_0] \sin \frac{\pi}{T_0} [(m-n)T_0 + u_m - u_n - \tau]}{2N\pi[(m-n)T_0 + u_m - u_n - \tau]}. \quad (153)$$

For purposes of calculation it is convenient to introduce the dummy variable z and write the equivalent form:

$$R_\sigma(\tau) = \int_0^{\pi/T_0} dz$$

$$\sum_{m=-N}^{N} \sum_{n=-N}^{N} \frac{T_0 R_s[(m-n)T_0] \cos [(m-n)T_0 + u_m - u_n - \tau]z}{2N\pi}. \quad (154)$$

Next we change the order of summation by letting $m - n = m'$ and dropping the prime after eliminating m to obtain

$$R_\sigma(\tau) = \int_0^{\pi/T_0} dz$$

$$\sum_{n=-N}^{N} \sum_{m=-N-n}^{n-N} \frac{T_0 R_s(mT_0) \cos (mT_0 + u_{m+n} - u_n - \tau)z}{2N\pi}. \quad (155)$$

The summation can be rearranged into the equivalent form

$$\sum_{m=-2N}^{0} \sum_{n=-N-m}^{N} + \sum_{m=1}^{2N} \sum_{n=-N}^{N-m},$$

so that the summation may be performed with respect to n first. For each value of m there are $2N + 1$ values of n. When N is large, summing over the values of the function

$$G(nT_0) = \frac{T_0 R_s(mT_0)}{\pi} \cos \{mT_0 + u[(m+n)T_0] - u(nT_0) - \tau\}z \quad (156)$$

for $2N + 1$ values of n with m fixed and then dividing by $2N$ approaches the procedure for calculating the average of $G(nT)$ over the u-ensemble.

To evaluate the ensemble average of $G(nT_0)$, we invoke a theorem expressed by equation (1.14) of Ref. 6, which states that, if $u(t)$ is a gaussian ensemble,

$$\text{av}\{\cos[au(t) + bu(t + \tau) + \beta]\}$$

$$= \exp\left[\frac{a^2 + b^2}{2} R_u(0) - abR_u(\tau)\right] \cos\beta. \qquad (157)$$

We shall assume from this point on that $u(t)$ is a gaussian ensemble. Then the above result fits our case if we set $t = nT_0$, $a = -z$, $b = z$, $\beta = (mT_0 - \tau)z$, $\tau = mT_0$. We then find

$$\text{av } G(nT_0) = \frac{T_0 R_s(mT_0)}{\pi} e^{-z^2[R_u(0) - R_u(mT_0)]} \cos(mT_0 - \tau)z, \qquad (158)$$

and

$$R_s(\tau) = \frac{T_0}{\pi} \sum_{m=-\infty}^{\infty} R_s(mT_0)$$

$$\int_0^{\pi/T_0} e^{-z^2[R_u(0) - R_u(mT_0)]} \cos[(\tau - mT_0)z] \, dz. \qquad (159)$$

The integral is now evaluated by writing the cosine in exponential form, completing the square in the exponent, and substituting variables to obtain the error function. The result is given as (129) of Section VI. We have replaced m by $-n$ and made use of the fact that the autocorrelation function is even.

We next calculate the power spectrum $w_s(f)$ by taking the Fourier transform of the approximate autocorrelation (130). The complete integrand has no singularities in τ and, by indenting the path of integration below each point $\tau = -nT_0$, we can separate the integral into component parts without singularities on the path. Denoting the depressed path by C, we evaluate

$$\int_C \frac{e^{-j2\pi f\tau} \sin\frac{\pi}{T_0}(\tau + nT_0)}{\pi(\tau + nT_0)} \, d\tau$$

$$= \int_C [e^{-j[2\pi f - (\pi/T_0)]\tau + jn\pi} - e^{-j[2\pi f + (\pi/T_0)]\tau - jn\pi}] \frac{d\tau}{2\pi j(\tau + nT_0)} \qquad (160)$$

$$= \begin{cases} 0; & |f| > \dfrac{1}{2T_0}, \\ e^{j2n\pi fT_0}; & |f| < \dfrac{1}{2T_0} \end{cases}$$

$$\int_c \frac{e^{-j2\pi f\tau} \cos \frac{\pi}{T_0}(\tau + nT_0)}{(\tau + nT_0)^2} d\tau$$

$$= \int_c [e^{-j[2\pi f - (\pi/T_0)]\tau + jn\pi} + e^{-j[2\pi f + (\pi/T_0)]\tau - jn\pi}] \frac{d\tau}{2(\tau + nT_0)^2} \quad (161)$$

$$= \begin{cases} 0; & |f| > \frac{1}{2T_0}, \\ \pi \left(2\pi f - \frac{\pi}{T_0}\right) e^{j2n\pi f T_0}; & |f| < \frac{1}{2T_0} \end{cases}$$

$$\int_c \frac{e^{-j2\pi f\tau} \sin \frac{\pi}{T_0}(\tau + nT_0)}{(\tau + nT_0)^3} d\tau$$

$$= \int_c [e^{-j[2\pi f - (\pi/T_0)]\tau + jn\pi} - e^{-j[2\pi f + (\pi/T_0)]\tau - jn\pi}] \frac{d\tau}{2j(\tau + nT_0)^3} \quad (162)$$

$$= \begin{cases} 0; & |f| > \frac{\pi}{2T_0} \\ -\frac{\pi}{2}\left(2\pi f - \frac{\pi}{T_0}\right)^2 e^{j2n\pi f T_0}; & |f| < \frac{1}{2T_0} \end{cases}$$

Then, for $|f| < 1/2T_0$; that is, $|f| < f_0$:

$$w_e(f) \doteq \sum_{n=-\infty}^{\infty} T_0 R_s(nT_0) \left[1 - \frac{\pi^2}{T_0^2} U_n^2 - 2U_n^2 \pi^2 \left(2f - \frac{1}{T_0}\right) \right.$$

$$\left. - 2\frac{T_0}{\pi}\frac{\pi}{2}\pi^2 \left(2f - \frac{1}{T_0}\right)^2 U_n^2 \right] e^{j2n\pi f T} \quad (163)$$

$$= T_0 \sum_{n=-\infty}^{\infty} R_s(nT_0)(1 - 4\pi^2 f^2 U_n^2) e^{j2n\pi f T}.$$

From (120), we note that $w_s(t)$ vanishes outside the range $-f_0$ to f_0,

$$R_s(nT_0) = \int_{-\infty}^{\infty} w_s(\nu) e^{j2\pi n T_0 \nu} d\nu, \quad (164)$$

and from the analogous relation for $R_u(t)$,

$$R_u(nT_0) = \int_{-\infty}^{\infty} w_u(\nu) e^{j2\pi n T_0 \nu} d\nu. \quad (165)$$

Application of the convolution theorem shows that

$$R_s(nT_0)R_u(nT_0) = \int_{-\infty}^{\infty} w_{s \cdot u}(\nu)e^{j2\pi n T_0 \nu} \, d\nu,$$ (166)

where

$$w_{s \cdot u}(\nu) = \int_{-\infty}^{\infty} w_s(\lambda)w_u(\lambda + \nu) \, d\lambda$$

$$= \int_{-\infty}^{\infty} w_s(\lambda + \nu)w_u(\lambda) \, d\lambda.$$ (167)

We substitute the above expressions for $R_s(nT_0)$ and $R_s(nT_0)R_u(nT_0)$ in (163) and then apply Poisson's summation formula,

$$\sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(z)e^{inz} \, dz = 2\pi \sum_{n=-\infty}^{\infty} \varphi(2n\pi),$$ (168)

to obtain, for $|f| < f_0$,

$$W_\sigma(f) \doteq [1 - 4\pi^2 f^2 R_u(0)] \sum_{n=-\infty}^{\infty} w_s\left(\frac{n}{T_0} - f\right)$$

$$+ 4\pi^2 f^2 \sum_{n=-\infty}^{\infty} w_{s \cdot u}\left(\frac{n}{T_0} - f\right).$$ (169)

Since both $w_s(f)$ and $w_u(f)$ have been assumed to vanish for $|f| > f_0 = 1/2T_0$, the only term of the first sum which falls in the signal band $-f_0$ to f_0 is that for $n = 0$, while in the second sum the signal-band contributors are $n = 0$, -1 and 1. The latter conclusion follows from the fact that the convolution of two functions limited to the same low-pass band is limited to a low-pass band twice as great. Also, noting that power spectra are even functions, we obtain (131).

REFERENCES

1. Rowe, H. E., this issue, pp. 1543–1598.
2. DeLange, O. E., this issue, pp. 1155–1486.
3. DeLange, O. E. and Pustelnyk, M., this issue, pp. 1487–1500.
4. Sunde, E. D., Self-Timing Regenerative Repeaters, B.S.T.J., **36**, July 1957, pp. 891–937.
5. Macfarlane, G. G., On the Energy Spectrum of an Almost Periodic Succession of Pulses, Proc. I.R.E., **37**, October 1949, pp. 1139–1143.
6. Bennett, W. R., Curtis, H. E. and Rice, S. O., Interchannel Interference in FM and PM Systems Under Noise Loading Conditions, B.S.T.J., **34**, May 1955, pp. 601–636.
7. Holbrook, D. B. and Dixon, J. T., Load Rating Theory for Multichannel Amplifiers, B.S.T.J., **18**, October 1939, pp. 624–644.
8. Feldman, C. B. and Bennett, W. R., Band Width and Transmission Performance, B.S.T.J., **28**, July 1949, pp. 490–595.

Timing in a Long Chain of Regenerative Binary Repeaters

By H. E. ROWE

The present paper studies some of the statistical properties of the random timing deviations, or position modulation of the signal pulses, in a long chain of regenerative binary repeaters. Random timing deviations of the output signal pulses result from input noise, tuning error, random timing deviations of the input signal pulses (introduced by preceding repeaters) and other sources at each repeater.

The power spectra and the total powers (mean square values) of the timing noise, spacing noise (random deviations in spacing of two consecutive pulses from an integral number of pulse periods) and alignment noise (random deviations in alignment between an input signal pulse and its corresponding timing pulse) caused by the input noise at each repeater are determined for a long chain of regenerative repeaters using either tuned circuit or locked oscillator timing filters. The effects of tuning error are studied for a chain of repeaters employing locked oscillator timing circuits; however, the present analysis does not treat the effects of tuning error in a chain of repeaters using tuned-circuit timing filters.

I. INTRODUCTION

Regenerative binary repeaters have recently been proposed for both baseband and carrier pulse code modulation systems,[1,2,3,4] in which the signal is represented by a binary pulse train. This type of repeater attempts to remove noise and other types of distortion from the incoming pulse train and to transmit a new signal which resembles the original as closely as possible. Noise and other system imperfections have two unwanted effects: (1) a certain number of errors occur at each repeater, i.e., a received pulse is transmitted as a space, or *vice versa*; (2) the signal pulses are no longer centered in equally spaced time slots but have a random position modulation, called timing noise. The present paper is concerned with some of the statistical properties of the random timing deviations in a system containing a long chain of regenerative repeaters.

A portion of a typical pulse train under ideal conditions is shown in Fig. 1, where, for purposes of illustration, the individual signal pulses have been made short enough so that they do not overlap; the curve represents either the amplitude of a baseband pulse train or the envelope of a carrier pulse train. The pulses are either "on" or "off", denoted by "1" and "0" respectively; all signal pulses present have a standard amplitude and identical shape, and are centered in equally spaced time slots.

An ideal regenerative repeater would sample the pulse train of Fig. 1 at the instants nT ($n = \cdots, -1, 0, 1, \cdots$). If the amplitude or envelope at each sample point is greater than the slicing level a new standard signal pulse is transmitted; if the amplitude or envelope is less than the slicing level no pulse is transmitted. If an additive gaussian noise is now present at the input to the regenerator there will be a certain number of errors, so that the output pulse train will no longer be identical to the input pulse train. To minimize the number of errors the slicing level should be set at one-half the peak pulse amplitude in a baseband system, a little greater than one-half the peak pulse envelope in a carrier system.[5] The error rate is then determined by the signal-to-noise ratio at the input to the regenerator.[5] In this way the effects of noise are completely eliminated, except in the relatively rare cases where the noise is large enough to cause an error.

These sampling and level-selecting operations may be performed by an idealized regenerator having the characteristics shown in Fig. 2, which approximate those of practical regenerators. In addition to the signal input and output, which may be either baseband or carrier pulses, the regenerator has an additional input for the timing or sampling pulses, which are baseband pulses. The input and output plotted in Fig. 2 repre-
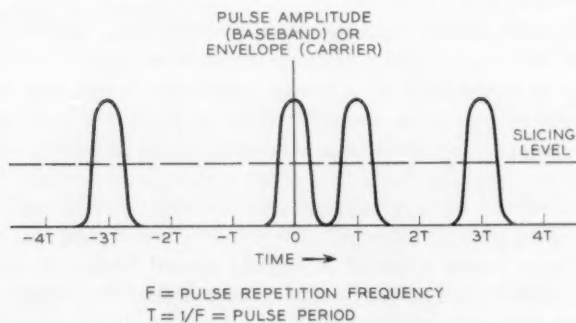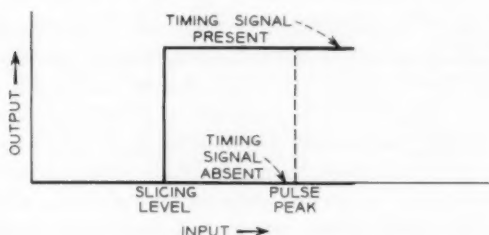

Fig. 1 — Binary pulse train.

Fig. 2 — Ideal regenerator characteristic.

sent amplitude or envelope for the baseband and carrier cases respectively. If the timing signal is absent, the regenerator output is zero for all input signal levels; if a timing pulse is present, the regenerator operates as an ideal slicer, with zero output for inputs less than the slicing level and a constant output for inputs greater than the slicing level. If the timing pulses are much shorter than the signal pulses, the regenerator output will consist of identical short baseband or carrier pulses, which may be transmitted through an appropriate filter to yield standard output signal pulses. Thus, the system will produce a standard output pulse each time a timing pulse occurs when the input signal has an amplitude or envelope greater than the slicing level. The position of this output pulse is determined only by the position of the timing pulse and not by the position of the input signal pulse; this type of response has been called "complete retiming".[3, 6]

In an ideal regenerative repeater with an input signal as shown in Fig. 1, the regenerator of Fig. 2 must be supplied with timing impulses occurring at the pulse repetition frequency and centered exactly at the sample points nT. However, in the self-timed repeaters studied here the timing pulses must be derived by the repeater itself from the signal, which will no longer be the ideal signal of Fig. 1 but will have added noise and random position modulation introduced by the preceding repeaters.

In systems employing complete retiming (such as the one discussed above), in which the timing pulse alone determines the position of the corresponding output signal pulse, the timing pulses may be derived only from the input signal. However, in systems employing partial retiming, in which the position of each output signal pulse depends on the position of both the corresponding timing and input signal pulses, the timing pulses may be derived from either the input or the output signal pulses.[3, 6] Repeaters with essentially complete retiming appear to be of greatest interest for microwave systems and so, for the present, we con-
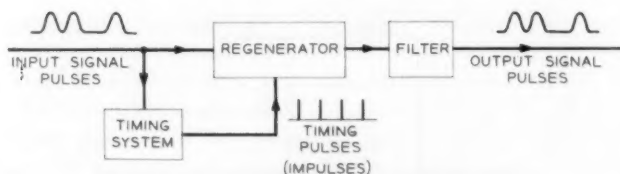
Fig. 3 — Regenerative binary repeater with timing from the input.

sider only the case of complete retiming. The analysis is easily extended to include partial retiming, with the timing pulses derived from either the input or the output.

The mathematical model chosen to represent a regenerative repeater is shown in Fig. 3. The regenerator characteristics are assumed to be those given in Fig. 2. The input and output signals are either baseband or carrier pulses; the timing signal is a train of baseband impulses, derived from the input pulse train. The filter converts the short pulses at the output of the regenerator into standard signal pulses for transmission to the next repeater.

The timing system in Fig. 3 contains a narrow band-pass filter tuned as close as possible to the pulse repetition frequency. This timing circuit is excited by baseband pulses derived from the input signal pulses. Its output, called the timing wave, is approximately a sine wave at the pulse repetition frequency, and may be considered to be a sine wave at this frequency with both random amplitude and phase modulation. The amplitude modulation results from the statistical nature of the signal pulse pattern and from the random noise introduced at the input to the repeater; the phase modulation is produced by the random noise at the input, the random variations in pulse position introduced by preceding repeaters in the chain, and the tuning error of the timing circuit. The timing circuit is followed by an ideal limiter which removes the amplitude modulation; the resulting waveform is then used to generate the timing pulses applied to the regenerator, for example, by producing a timing pulse at the instant of each negative- (or positive-) going zero crossing of this waveform, so that the phase of the timing wave determines the position of the timing pulses.

The baseband driving pulses for the timing circuit may be obtained in various ways. The simplest method is to drive the timing circuit with the signal pulses themselves in a baseband system or with their rectified envelope (obtained by a linear envelope detector) in a carrier system. However, it may be advantageous to first pass the baseband signal pulses

through a nonlinear device such as a peak amplifier or a square-law rectifier to suppress the low-level noise in the absence of signal pulses;[2] similarly, a square-law rectifier might be used for the same purpose in a carrier system. Alternately, a standard baseband driving pulse for the timing circuit might be generated whenever the amplitude or envelope of the baseband or carrier signals passes through a critical level (approximately half the peak) in the ascending direction.[7] The noise at the input of each repeater will affect the timing performance of these different systems in different ways.

In order to simplify the analysis of the timing behavior of a long chain of repeaters and to permit the application of the results to systems employing different types of repeaters, such as those discussed above, the following general assumptions will be made:

1. The reception of each signal pulse is assumed to initiate an independent transient in the timing circuit. This requirement is obviously satisfied in any baseband or carrier system in which signal pulses in adjacent time slots do not overlap. It will remain satisfied if adjacent signal pulses do overlap only under special conditions, e.g., (a) in a baseband system, if the received signal pulses drive the timing circuit directly, (b) in a carrier system with a linear envelope detector and coherent carrier phase between adjacent signal pulses.

2. The timing filter is assumed to be a simple resonant circuit, characterized by its Q. Its natural resonant frequency would ideally be made equal to the pulse repetition frequency, but in practice there will be a small tuning error. A related problem in which an idealized locked oscillator is used to generate the timing wave will also be considered.

3. The input noise at each repeater causes random timing deviations of the output signal pulses, in addition to those present from other sources. This real input noise may be replaced by adding an equivalent fictitious position modulation or timing noise to the input signal pulses, such that the random timing deviations at the output of the repeater remain the same. This permits the effects of the input noise to be treated in the same manner as the effects of the random position modulation introduced by the preceding repeaters. For simplicity, we assume that the equivalent timing deviations added to the different input signal pulses are statistically independent, the mean square value of the added timing noise being given. This assumption is plausible, since the real input noise will have a bandwidth comparable to that of the signal. A detailed analysis of a particular repeater is required, of course, to establish this equivalence rigorously; the rms value of the equivalent timing noise may depend on the average number of pulses present as well as on

the noise level at the repeater input. While no attempt will be made to give detailed consideration to any of the various types of repeaters discussed above, the dependence of the equivalent added timing noise on the pulse pattern will be discussed briefly for several cases of interest.

4. In order to permit the treatment of the timing behavior of a chain of repeaters in a simple way, approximations are made which linearize the analysis. It is then strictly valid only when the equivalent timing noise added at the input of each repeater is vanishingly small. Although the effects of these approximations are not known in detail, it seems plausible that the analysis remains valid in the practical case where the input noise at each repeater is small.

5. The input signal-to-noise ratio must be moderately high in a satisfactory system so that very few errors are made in recognizing pulses and spaces; these errors are neglected in the present timing analysis.

Consider first the case where the timing filters are simple tuned circuits. Even with the above approximations a rigorous solution for the timing behavior of a chain of repeaters has been obtained only under special conditions. In particular, the effects of the input noise may be determined if the tuning error of every repeater is zero, for only the following signal pulse patterns: (1) all pulses present; (2) every Mth pulse present; (3) any general periodic pulse pattern. In the first two cases the analysis is straightforward, but in the third case the complexity increases with the complexity of the pulse pattern and becomes somewhat prohibitive for all but the simpler periodic pulse patterns. The most interesting case — a random pulse pattern — has been treated in a simple way only by making a further approximation, in which the variation of the timing wave amplitude is neglected at an appropriate point in the analysis.[7] In general, no accurate estimate is avialable for the error introduced by this approximation. It gives accurate results in two cases which can be solved by other methods. These are: (1) a chain of repeaters with a periodic pulse pattern containing two pulses located in arbitrary positions, as discussed above; (2) a single repeater with a random pulse pattern, which has been treated by W. R. Bennett in a different way.[7]

Other problems concerning a chain of repeaters employing tuned circuits as timing filters, which may well be of greater practical importance than the effects of input noise, have not been treated in the present analysis. The first of these is the effect of random tuning errors at the different repeaters. This problem has been treated for a single repeater by W. R. Bennett.[7] The second problem is the effect of the finite width of the pulses used to drive the timing circuit, which will be shown to add an identical timing noise at the output of each repeater and to be quite

similar to the effects of amplitude-to-phase conversion in the limiter in the timing system.[2,8] Since the added timing noise is identical at each repeater, this type of disturbance must be treated differently from independent random disturbances added at the different repeaters, and is potentially more serious.

As stated above, in order to obtain an approximate analysis for a chain of repeaters with zero tuning error and a random pulse pattern we modify the timing response of each repeater by neglecting the variation in timing wave amplitude. Alternately, we may ask what kind of system is exactly described by the modified equations. It turns out that a somewhat idealized locked oscillator, synchronized by the signal pulses or their envelope, corresponds exactly to this analysis, within the other approximations described above. For a chain of repeaters using locked oscillators as timing circuits, both the effects of input noise and the effects of tuning error can be treated for arbitrary or random pulse patterns without further approximations. Also, finite pulse width introduces no additional timing noise in this case. Thus, the analysis for locked oscillators is considerably more tractable than is that for tuned circuits.

The statistical properties of three different quantities are of interest in studying the timing behavior of a repeater chain. These are:

1. *Timing noise*, or random deviations of the signal pulses from equally spaced time slots, will cause a random delay modulation of the original signal, unless special precautions are taken in the final decoding of the PCM signal. This random delay modulation will, for example, produce crosstalk in a frequency-division multiplex signal,[2,7] and will degrade other types of signals in different ways.[2]

2. *Spacing noise*, or the random deviations in the spacing of two consecutive signal pulses (which do not necessarily occupy adjacent time slots) from an integral number of pulse periods could cause pulses occupying adjacent time slots to interfere with each other. This would degrade the performance of the regenerator and cause an increased number of errors, i.e., a received pulse transmitted as a space, or *vice versa*. If too large, it could also cause some received pulses to be assigned to incorrect time slots, with resulting errors in the decoding of the PCM signal.

3. *Alignment noise*, or the random deviations in alignment between an input signal pulse and the corresponding timing pulse, can also degrade the performance of the regenerator and increase the number of errors, since for optimum margin against noise the timing or sampling pulse should fall exactly at the center of the corresponding signal pulse.

These three quantities are defined in Fig. 4, which will be discussed in greater detail below.

The present analysis determines the power spectra and the total powers (mean square values) for these three types of deviations for a long chain of repeaters, under the conditions discussed above. The way in which these disturbances vary along a repeater chain is obviously of great importance. It is found that, while the low-frequency timing noise does grow along the chain, the spacing and alignment noise spectra approach limiting values rather quickly, as might be expected on physical grounds. Since the timing, spacing and alignment deviations are discrete functions, i.e., defined only for integral values of their agrument, their power spectra are defined somewhat differently than in the usual case of continuous functions; a brief description of the Fourier analysis of such discrete functions is included.

Related studies of the timing deviations in a repeater chain have been given by DeLange[2] and Sunde,[3] using different methods than those employed here. The present analysis follows closely the approach first used by J. R. Pierce in determining the timing deviations in a single repeater,[6] extending the analysis to a chain of repeaters and to include the spacing and alignment deviations.

II. THE TIMING RESPONSE OF A SINGLE REPEATER

In this section we relate the timing deviations of the output pulses to the timing deviations of the input pulses and to the pulse pattern for a single repeater. As discussed above, we consider both tuned-circuit and locked-oscillator timing circuits. In order to study the response of a repeater to the timing deviations of the input pulses it is necessary to consider only the simplest case of impulse excitation of the timing circuit. Finite pulse width may introduce additional output timing deviations for tuned circuit timing filters but not for idealized locked oscillators; excitation of a tuned circuit by raised cosine pulses is therefore also considered.

2.1 *Tuned Circuit Excited by Impulses*[6]

Assume the timing circuit to be a parallel resonant inductance L, capacitance C and resistance R. The impulse response of this circuit is given by the real part of the complex impulse response $H(t)$:

$$H(t) = \frac{1}{C}\left(1 + \frac{j}{2Q}\right) e^{-(\pi/Q)f_0 t} e^{+j2\pi f_0 t}; \quad t > 0, \tag{1}$$

where

$$\omega_0 = 2\pi f_0 = \sqrt{\frac{1}{LC} - \frac{1}{(2RC)^2}}, \tag{2}$$

$$Q = \omega_0 RC, \qquad \frac{\omega_0}{2Q} = \frac{1}{2RC}.$$

In (1) and in all subsequent complex expressions the real part is implied.

The natural resonant frequency f_0 (as distinguished from the steady-state resonant frequency $f_r = 1/2\pi\sqrt{LC}$) will be made as close as possible to the pulse repetition frequency F. Therefore we set

$$f_0 = F + \delta f = F\left(1 + \frac{\delta f}{F}\right), \tag{3}$$

where

$\quad F = $ pulse repetition frequency,

$\quad \delta f = $ tuning error of the timing circuit,

$\quad F = \dfrac{1}{T}$, T being the pulse period (Fig. 1),

$\quad \dfrac{\delta f}{F} \ll 1$.

Substituting into (1), the response to a unit impulse at the time t_0 is equal to the real part of

$$H(t - t_0) = \frac{1}{C}\left(1 + \frac{j}{2Q}\right) e^{-(\pi/Q)F(1+\delta f/F)(t-t_0)} e^{j2\pi F(1+\delta f/F)(t-t_0)}; \tag{4}$$

$$t - t_0 > 0.$$

The pulse train driving the tuned circuit is given by

$$g(t) = \sum_{n=-\infty}^{\infty} a_n \delta(t - t_n), \tag{5}$$

where $\delta(t - t_n)$ is a unit impulse occurring at t_n, the time of arrival of the signal pulse corresponding to the nth time slot. If the nth time slot contains a pulse $a_n = 1$; if this time slot is vacant $a_n = 0$. In the absence of timing noise the signal pulses would be centered in their corresponding time slots, so that for the pulse corresponding to the nth time slot

$$t_n = nT. \tag{6}$$

If the deviation in the position of this pulse is δt_n, then the time of ar-

rival becomes

$$t_n = nT + \delta t_n .\tag{7}$$

It is convenient to normalize the timing deviation with respect to the pulse period T. Thus, let

$$\epsilon^i(n) = \frac{\delta t_n}{T},\tag{8}$$

where $\epsilon^i(n)$ is the normalized timing deviation of the signal pulse in the nth time slot at the input to the repeater, referred to simply as the timing deviation when no confusion will arise. Then the time of arrival of this pulse in (7) becomes

$$t_n = nT + \epsilon^i(n)T,$$
$$Ft_n = n + \epsilon^i(n).\tag{9}$$

From (4), (5) and (9) the response of the timing circuit to all of the signal pulses up to and including the pulse in the mth time slot is equal to the real part of

$$G(t) = \frac{1}{C}\left(1 + \frac{j}{2Q}\right)\sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(1+\delta f/F)[Ft-n-\epsilon^i(n)]}$$
$$\cdot e^{j2\pi(1+\delta f/F)[Ft-n-\epsilon^i(n)]}.\tag{10}$$

This expression gives the output of the timing circuit for values of t lying between the arrival times of the pulse in the mth time slot and the next pulse that is present in the signal. Formally,

$$t_m < t < t_k,$$
$$m + \epsilon^i(m) < Ft < k + \epsilon^i(k),$$
$$a_n = \begin{cases} 1; & n = m, k \\ 0; & m < n < k. \end{cases}\tag{11}$$

The response of the timing circuit given in (10) may be written in the form of a carrier at the pulse repetition frequency F with both amplitude and phase modulation:

$$G(t) = \frac{1}{C}\left(1 + \frac{j}{2Q}\right)A(t)e^{j2\pi Ft}; \qquad A(t) = a(t)e^{j\varphi(t)}.\tag{12}$$

In this equation, $A(t)$ is a complex function of time whose magnitude $|A(t)| = a(t)$ equals the normalized amplitude of the timing wave and

whose angle $\angle A(t) = \varphi(t)$ equals the phase deviation of the timing wave. For convenience, we have chosen to remove the constant factor $(1/C)\ (1 + j/2Q)$ in defining $A(t)$. From (10),

$$A(t) = \sum_{n=-\infty}^{\infty} a_n e^{-(\pi/Q)(1+\delta f/F)[Ft-n-\epsilon^i(n)]} e^{j2\pi\{(1+\delta f/F)[Ft-n-\epsilon^i(n)]-Ft\}}, \quad (13)$$

keeping in mind the restrictions of (11).

Equation (13) shows the way in which the amplitude and phase of the timing wave vary with time when timing noise is present. The amplitude decreases exponentially between signal pulses, increasing abruptly at the instant a pulse is received. In the absence of tuning error the phase is constant between signal pulses, changing abruptly when a pulse is received. If the tuning error is not zero, the phase has, in addition, a small constant linear variation between signal pulses. If the tuning error δf and the input timing deviations $\epsilon^i(n)$ are both equal to zero, we have from (13)

$$\angle A(t) = \varphi(t) = 0. \quad (14)$$

It is convenient to assume that under these conditions the delays in the repeater of Fig. 3 have been adjusted so that the timing pulses supplied to the regenerator are properly aligned with the input signal pulses, i.e., occur at the instants nT in Fig. 1. The quantity $\varphi(t)$ then gives a true measure of the timing deviations of the timing pulses.

We now assume that the phase of the timing wave at the instant immediately following the reception of a signal pulse determines the timing deviation of the corresponding timing pulse. This, of course, may not be strictly true. For example, the timing pulse might be generated at the next negative-going zero crossing of the timing wave, as discussed in the introduction. This will occur approximately $T/4$ seconds after the arrival of the signal pulse because, as shown in the subsequent analysis, in any satisfactory system the alignment error (defined in the introduction) remains small, and consequently the maxima of the timing wave occur close to the driving pulses for the timing circuit. However, for zero tuning error the timing wave phase $\varphi(t)$ is constant between signal pulses and consequently may be evaluated equally well at any time during the pulse period T following the arrival of a signal pulse. Even if the tuning error is not zero $\varphi(t)$ changes very slowly during the interval T, so that a small error in the time at which it is evaluated becomes unimportant; evaluating the phase just after the arrival of a signal pulse rather than $T/4$ seconds later will cause a very small error in the dc value of the output timing deviation, but will have no other effect on the analysis.

Consequently, evaluating (13) at the time

$$Ft = Ft_m = m + \overset{i}{\epsilon}(m) \tag{15}$$

we obtain

$$A(t_m)^{\sim}_{-} = e^{-j2\pi \overset{i}{\epsilon}(m)} \sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(1+\delta f/F)[(m-n)+\overset{i}{\epsilon}(m)-\overset{i}{\epsilon}(n)]}$$
$$\cdot e^{j2\pi\{\overset{i}{\epsilon}(m)-\overset{i}{\epsilon}(n)+(\delta f/F)[(m-n)+\overset{i}{\epsilon}(m)-\overset{i}{\epsilon}(n)]\}}. \tag{16}$$

Denoting the normalized output timing deviation of the repeater by $\overset{o}{\epsilon}(m)$, defined as in (8), we have for the case of complete retiming, where the timing deviation of each output signal pulse is identical to the timing deviation of the corresponding timing pulse,

$$\overset{o}{\epsilon}(m) = -\frac{\varphi(t_m)}{2\pi}; \qquad \varphi(t_m) = \angle A(t_m), \tag{17}$$

where $A(t_m)$ is given by (16). The minus sign in (17) occurs because the timing pulse will occur too soon for $\varphi(t)$ positive, corresponding to a negative timing deviation according to (7) and (8).

Equations (16) and (17) determine the general relation between the input and output timing deviations for a single repeater. The output timing deviation is seen to be a rather complicated function of the past input timing deviations. In order to make possible a reasonably simple analysis that can readily be extended to a chain of repeaters, certain restrictions will be imposed that linearize the relations between input and output timing deviations.

Equation (16) gives $A(t_m)$ as the sum of an infinite number of vectors, each associated with one of the past input signal pulses. The amplitudes of these vectors become exponentially smaller the farther back in time the corresponding signal pulse occurred. If the angles between those vectors which give the essential contribution to the sum are small enough so that the sine of an angle is approximately equal to the angle, then the vector summation is quite easily accomplished. Taking the vector corresponding to the present input pulse ($n = m$) as a reference, the in-phase component is approximately the sum of the magnitudes of all of the vectors, the quadrature component is the sum of the magnitudes times the relative angles. In order for these approximations to be valid we must have

$$\left.\begin{array}{c} \left|\dfrac{\delta f}{F}(m-n)\right| \ll 1 \\[2mm] |\overset{i}{\epsilon}(m) - \overset{i}{\epsilon}(n)| \ll 1 \end{array}\right\} \text{ for all significant } n. \tag{18}$$

The significant terms in the summation are for values of n bounded by

$$n_0 < n < m,$$

where

$$(m - n_0) \gg \frac{Q}{\pi}. \tag{19}$$

The first condition of (18) thus becomes

$$\left| \frac{\delta f}{F} \right| \ll \frac{\pi}{Q}. \tag{20}$$

There are two different ways in which the second condition of (18) may be satisfied. Either the input timing deviation must be small compared to one or it must change so slowly that the difference in the deviations of pulses separated by Q/π pulse periods will be small compared to one. Thus,

$$| \epsilon^i(n) | \ll 1$$

or

$$| \Delta\epsilon^i(n) | = | \epsilon^i(n + 1) - \epsilon^i(n) | \ll \frac{\pi}{Q}, \tag{21}$$

where $\Delta\epsilon^i(n)$ represents the first forward difference of $\epsilon^i(n)$.

Subject to (20) and one of the conditions of (21) we have, from (16) and (17), the normalized timing wave amplitude A_m and the normalized output timing deviation $\epsilon^0(m)$ (with complete retiming):

$$A_m = | A(t_m) | = \sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(m-n)}, \tag{22}$$

$$\epsilon^0(m) = \frac{\displaystyle\sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(m-n)} \left[\epsilon^i(n) - \frac{\delta f}{F} (m - n) \right]}{A_m}. \tag{23}$$

Alternately, with the summations rewritten, (22) and (23) become

$$A_n = \sum_{k=0}^{\infty} a_{n-k} e^{-(\pi/Q)k}, \tag{24}$$

$$\epsilon^0(n) = \frac{\displaystyle\sum_{k=0}^{\infty} a_{n-k} e^{-(\pi/Q)k} \left[\epsilon^i(n - k) - \frac{\delta f}{F} k \right]}{A_n}. \tag{25}$$

For zero tuning error, (22) and (23) or (24) and (25) give the output tim-

ing deviation as a linear function of past input timing deviations for a repeater with complete retiming, subject to the restriction of (20) and (21). It is usually implied that $a_m = 1$ in (23) or $a_n = 1$ in (25), since the output timing deviation has meaning only for values of its argument corresponding to time slots that contain signal pulses. The timing wave phase corresponding to vacant time slots will not normally be of interest (except insofar as there may be a preferred time to examine a vacant time slot to determine that no signal pulse is present).

These equations show that, for $\delta f = 0$, each repeater may be considered a linear transducer to the timing deviations. In general, the equivalent transducer will be time-varying, since the timing wave amplitude A_n will vary from pulse to pulse; the pulse pattern enters explicitly into the analysis, both in determining A_n and through the a's in the numerator of (23) or (25). However, if all pulses are present or if the pulse pattern is periodic with every Mth pulse present (all pulses present correspond to the special case $M = 1$), A_n is constant at every signal pulse. This is the only case in which the repeater acts as a strictly invariant linear transducer, permitting a simple analysis of a chain of repeaters.

In dealing with general pulse patterns it is convenient to define a new "primed" independent variable that numbers the signal pulses consecutively, rather than the time slots. Fig. 4(a) shows a portion of a typical pulse train; consecutive time slots are denoted by the variable n, consecutive pulses by n'. A quantity regarded as a function of n' or any other primed independent variable will be distinguished by the symbol $\sim$; functions of n or any other unprimed independent variable will be written as before. For any given pulse pattern n is a function of n', which of course differs for each different pulse pattern, and *vice versa*. As shown in Fig. 4, the number of vacant time slots between the $(n' - 1)$th and the n'th (consecutive) pulses is defined as $b_{n'}$. Referring to Fig. 4, the following examples illustrate this notation:

$\epsilon(n)$ = timing deviation of pulse corresponding to the nth time slot,

$\tilde{\epsilon}(n')$ = timing deviation of the n'th pulse,

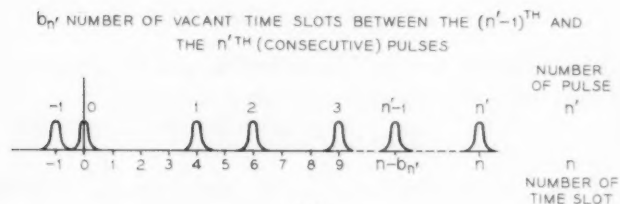A_n = timing wave amplitude corresponding to the nth time slot,

$\tilde{A}_{n'}$ = timing wave amplitude corresponding to the n'th pulse,

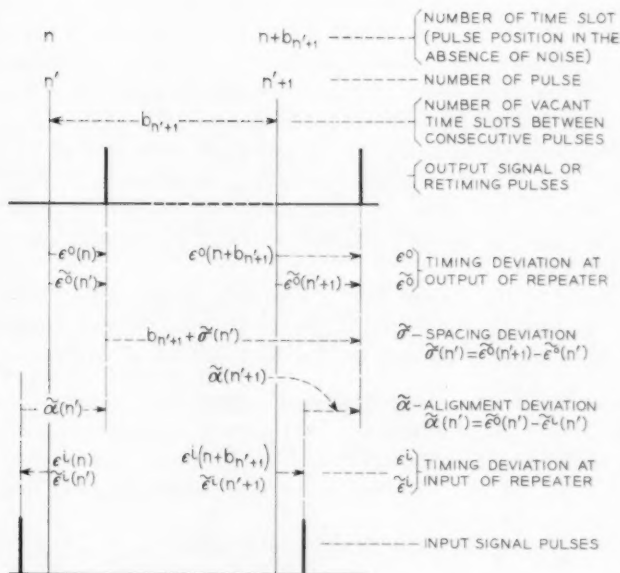$$\tilde{\epsilon}(n' - 1) = \epsilon(n - b_{n'}), \qquad \tilde{\epsilon}(n') = \epsilon(n),$$
$$\tilde{A}_{n'-1} = A_{n-b_{n'}}, \qquad \tilde{A}_{n'} = A_n . \tag{26}$$

Thus, for the particular pulse pattern of Fig. 4(a) we have, for example:

$$\tilde{\epsilon}(0) = \epsilon(0); \qquad \tilde{\epsilon}(2) = \epsilon(6); \qquad \tilde{A}_3 = A_9 .$$

$b_{n'}$ NUMBER OF VACANT TIME SLOTS BETWEEN THE $(n'-1)^{TH}$ AND
THE n'^{TH} (CONSECUTIVE) PULSES

NUMBER
OF PULSE
n'

-1 0   1   2   3 $n'-1$ n'

-1 0 1 2 3 4 5 6 7 8 9 $n-b_{n'}$ n

n
NUMBER OF
TIME SLOT

(a)

n $n+b_{n'+1}$ - - - - - - {NUMBER OF TIME SLOT
                                                  (PULSE POSITION IN THE
                                                  ABSENCE OF NOISE)

n' $n'+1$ - - - - - - NUMBER OF PULSE

$b_{n'+1}$ {NUMBER OF VACANT
                                                  TIME SLOTS BETWEEN
                                                  CONSECUTIVE PULSES

                                                 {OUTPUT SIGNAL OR
                                                  RETIMING PULSES

$\epsilon^0(n)$ $\epsilon^0(n+b_{n'+1})$ ϵ^0} TIMING DEVIATION AT
$\tilde{\epsilon}^0(n')$ $\tilde{\epsilon}^0(n'+1)$ $\tilde{\epsilon}^0$} OUTPUT OF REPEATER

$b_{n'+1}+\tilde{\sigma}(n')$ $\tilde{\sigma}$— SPACING DEVIATION
                                                  $\tilde{\sigma}(n')=\tilde{\epsilon}^0(n'+1)-\tilde{\epsilon}^0(n')$

                       $\tilde{\alpha}(n'+1)$ $\tilde{\alpha}$—ALIGNMENT DEVIATION
                                                  $\tilde{\alpha}(n')=\tilde{\epsilon}^0(n')-\tilde{\epsilon}^L(n')$
$\tilde{\alpha}(n')$

$\epsilon^L(n)$ $\epsilon^L(n+b_{n'+1})$ ϵ^L} TIMING DEVIATION AT
$\tilde{\epsilon}^L(n')$ $\tilde{\epsilon}^L(n'+1)$ $\tilde{\epsilon}^L$} INPUT OF REPEATER

                                        - - - - - INPUT SIGNAL PULSES

(b)

Fig. 4 — Definition of timing, spacing and alignment deviations for a repeater with complete retiming: (a) general pulse pattern, in absence of timing noise; (b) timing, spacing and alignment deviations.

For all pulses present $n' = n$ and $b_{n'} = 1$, and the two notations coincide.

The timing, spacing and alignment deviations defined in Section I are illustrated in Fig. 4(b). The timing deviation of the input and the timing and output pulses have been defined above. The spacing deviation is given by

$$\tilde{\sigma}(n') = \tilde{\epsilon}^o(n' + 1) - \tilde{\epsilon}^o(n') = \Delta\tilde{\epsilon}^o(n'), \tag{27}$$

where Δ indicates the first forward difference, and the alignment deviation for a repeater with complete retiming is given by

$$\tilde{\alpha}(n') = \tilde{\epsilon}^o(n') - \tilde{\epsilon}^i(n'), \tag{28}$$

recalling that, with complete retiming, the timing deviations of the timing and the output signal pulses are equal.

The choice of n' as the independent variable in the analysis rather than the original independent variable n is more natural in several ways. As mentioned above, we are usually interested in the timing wave phase only for time slots that contain pulses. The definition of spacing deviation in (27) is much easier to work with than it would be if it were in terms of the original variable n. As will appear below, this change of independent variable facilitates the approximate treatment of a chain of repeaters with random or general periodic pulse patterns. While it would present some complications in dealing with quantities that must be expressed in real time, no such difficulties arise in the present treatment of a repeater chain.

Returning to the special case of a periodic pulse pattern with every Mth pulse present, the results of (21) to (25) may be written very conveniently in terms of the new independent variable n'. We have

$$a_n = \begin{cases} 1; & n = Mn' \\ 0; & \text{otherwise} \end{cases} \tag{29}$$

and

$$n = Mn'; \qquad k = Mk'. \tag{30}$$

The timing wave amplitude takes on the constant value A at every signal pulse; from (24),

$$A = A_{Mn'} = \tilde{A}_{n'} = \sum_{k'=0}^{\infty} e^{-(\pi M/Q)k'}, \qquad A = \frac{1}{1 - e^{-\pi M/Q}}. \tag{31}$$

The output timing deviation in (25) becomes

$$\begin{aligned}
\tilde{\epsilon}^o(n') &= (1 - e^{-\pi M/Q})\left[\sum_{k'=0}^{\infty} \tilde{\epsilon}^i(n' - k')e^{-(\pi M/Q)k'} - \frac{\delta f}{F}M\sum_{k'=0}^{\infty} k'e^{-(\pi M/Q)k'}\right] \\
&= (1 - e^{-\pi M/Q})\sum_{k'=0}^{\infty} \tilde{\epsilon}^i(n' - k')e^{-(\pi M/Q)k'} - \frac{\delta f}{F}\frac{Me^{-\pi M/Q}}{1 - e^{-\pi M/Q}}.
\end{aligned} \tag{32}$$

For the high-Q case we may set

$$\frac{1}{A} = (1 - e^{-\pi M/Q}) \approx \frac{\pi M}{Q}; \qquad Q \gg \pi M \qquad (33)$$

in (31) and (32), so that (32) becomes

$$\bar{\epsilon}^\circ(n') = \frac{\pi M}{Q} \sum_{k'=0}^{\infty} \bar{\epsilon}^i(n' - k') e^{-(\pi M/Q)k'} - \frac{\delta f}{F} \frac{Q}{\pi}; \qquad Q \gg \pi M. \quad (34)$$

The second term of (32) and (34) represents a constant timing deviation introduced by the tuning error; the first term shows that the repeater is a strictly invariant linear transducer to the input timing deviations for this special case.

In a chain of repeaters the timing wave amplitude A_n is identical at each repeater, since the pulse pattern is the same at each repeater. Thus, for random or general periodic pulse patterns we must treat the repeaters as identical linear time-varying transducers to the timing deviations. Now, if Q is large, the variation in A_n will be small;[7] we can make use of this fact to treat the repeaters as approximately invariant transducers. Equations (23) and (25) may be written in an alternate form, with n' as the independent variable, that has a simple physical interpretation and proves useful in making the further approximation that permits the analysis of a chain of repeaters with random or general periodic pulse patterns, for zero tuning error.

We now examine the change in the output timing deviation that occurs during the interval between two consecutive signal pulses. Making use of (26), we have, from (24) and (25), the first backward difference of the output timing deviation as a function of n':

$$\bar{\epsilon}^\circ(n') - \bar{\epsilon}^\circ(n' - 1) = \frac{1}{\bar{A}_{n'}} \left[\bar{\epsilon}^i(n') - \bar{\epsilon}^\circ(n' - 1) + \frac{\delta f}{F} b_{n'} \right] - \frac{\delta f}{F} b_{n'} . \quad (35)$$

The timing wave amplitude is given by (24), as before. Referring to Fig. 5, we may interpret this relation as follows: Immediately after the reception of the $(n' - 1)$th pulse the timing wave amplitude is $\bar{A}_{n'-1}$ and the output timing deviation is $\bar{\epsilon}^\circ(n' - 1)$, with corresponding timing wave phase $-2\pi\bar{\epsilon}^\circ(n' - 1)$. During the interval of $b_{n'}$ pulse periods from the arrival of the $(n' - 1)$th pulse until just before the arrival of the n'th pulse the timing wave amplitude will decay exponentially to the value $e^{-(\pi/Q)b_{n'}}\bar{A}_{n'-1}$ and the timing wave phase will advance linearly by an angle $2\pi(\delta f/F)b_{n'}$, so that the resultant phase is

$$-2\pi[\bar{\epsilon}^\circ(n' - 1) - (\delta f/F)b_{n'}],$$

as illustrated by the top vector coming from the origin in Fig. 5. The n'th pulse, with timing deviation $\tilde{\epsilon}^i(n')$, initiates a corresponding unit vector of phase $-2\pi\tilde{\epsilon}^i(n')$, which adds to the vector representing the previous value of the timing wave to give a resultant timing wave vector of amplitude $\tilde{A}_{n'}$, phase $-2\pi\tilde{\epsilon}^o(n')$. Since the relevant angles are small, the result given in (35) may be readily obtained.

Equations (35) and (25) are completely equivalent statements describing the timing response of a single repeater using a resonant circuit to generate the timing wave. They are derivable from each other without any approximation and are both subject to the restrictions given in (20) and (21). They provide somewhat different representations of the repeater as a linear time-varying transducer to the timing deviation. For the special case of a periodic pulse pattern with every Mth pulse present and $\tilde{A}_{n'}$ constant, given by (31), in which the repeater becomes a linear invariant transducer, (35) is equivalent to (32). We shall show in Section 2.2 that (35) is in the appropriate form for the approximate treatment of a chain of repeaters with zero tuning error for random and general periodic pulse patterns, by neglecting the variation of the timing wave amplitude.
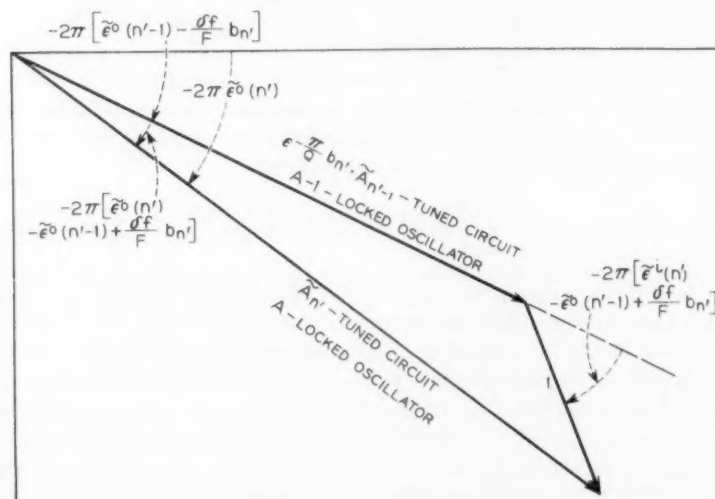
Fig. 5 — Change in output timing deviation during the interval between two consecutive signal pulses.

2.2 *Approximate Timing Response for Random and General Periodic Pulse Patterns — Tuned Circuit Timing Filter, Zero Tuning Error.*

W. R. Bennett has pointed out that, for a high enough Q, the variation of timing wave amplitude in a tuned circuit will be very small, and has used this fact in his treatment of a single repeater.[7] Making use of this fact, we show how a repeater using a tuned circuit timing filter with zero tuning error may be treated as an approximately invariant linear transducer to the timing deviations.

We might first think that we could set $\tilde{A}_{n'}$ equal to its average value in (25), but this does not yield a satisfactory approximation to the behavior of a repeater. As pointed out in the discussion following (17), the numerator of (25) is simply the quadrature component and the denominator is the in-phase component of the timing wave, referred to the transient started by the present input pulse. For zero tuning error, both numerator and denominator decay exponentially between signal pulses, so that the timing wave phase remains constant; this is apparent either on physical grounds or on examination of (24) and (25). If we arbitrarily set $\tilde{A}_{n'}$ constant in (25), we should obtain an exponentially decreasing phase deviation between signal pulses, rather than a constant deviation. Further, for zero tuning error a constant dc input timing deviation produces an equal constant output timing deviation for any arbitrary pulse pattern. If, however, we set $\tilde{A}_{n'}$ constant in (25) the output timing deviation fluctuates about its average value, so that it would now contain false ac components. Since it turns out that the timing deviation in a repeater chain consists principally of low-frequency components, an approximate treatment of a single repeater should yield accurate results for dc and slowly varying input timing deviations.

The correct approach to this problem has been indicated by J. L. Kelly, Jr. If the tuning error is zero, a useful approximation to the behavior of a repeater is obtained by setting the timing wave amplitude equal to its average value in (35). Thus, setting $\delta f = 0$, we have

$$\bar{\epsilon}^o(n') - \bar{\epsilon}^o(n' - 1) = \frac{1}{A} [\bar{\epsilon}^i(n') - \bar{\epsilon}^o(n' - 1)], \qquad A = \langle \tilde{A}_{n'} \rangle, \quad (36)$$

where the brackets $\langle \ \rangle$ indicate an ensemble average for a random pulse pattern, and A_n is given by (24). Referring to Fig. 5 and to the discussion following (35), in this approximation the phase deviation of the timing wave remains constant between signal pulses, as it should. A dc input timing deviation produces an equal dc output timing deviation, with no false ac components; the error in this approximation should be

small for the low-frequency components that form the major part of the timing deviation.

Equation (36) describes a repeater with a tuned circuit timing filter with zero tuning error as a linear invariant transducer to the timing deviation, greatly simplifying the analysis of a chain of repeaters. For random and general periodic pulse patterns the variation in the timing wave amplitude will modify the timing behavior somewhat; for periodic pulse patterns with every Mth pulse present (including all pulses present) (36) is exact. The average timing wave amplitude A and its standard deviation are determined in Section 2.4; A is proportional to the average number of pulses present and to the tuned-circuit Q. We shall show in Section 2.3 that (36) gives an exact description of a locked oscillator with zero tuning error for any arbitrary pulse pattern.

2.3 Locked Oscillator Timing Circuit

In the preceding section the approximate timing response of a repeater with zero tuning error was found by setting $\tilde{A}_{n'}$ equal to its average value in (35). It is natural to ask what kind of system is described exactly by (35) with $\tilde{A}_{n'}$ equal to a constant A (δf is now not necessarily equal to zero).

Such a system may be regarded as an idealized locked oscillator. The discussion of Fig. 5 immediately following (35) remains applicable except that the amplitude of the timing wave no longer decays between signal pulses but has a constant length ($A - 1$). The unit vector initiated by the n'th pulse adds to the vector representing the previous value of the timing wave, causing an abrupt change in its phase as before and increasing its length momentarily to A. However, we now assume that the timing wave amplitude returns to its constant value ($A - 1$) before the next signal pulse arrives. In a practical locked oscillator with a weak synchronizing signal, the transients started by the signal pulses will be so much smaller than the steady-state oscillator wave that the timing wave amplitude will be essentially constant, independent of the pulse pattern.

Setting $\tilde{A}_{n'} = A$ in (35), we have an exact expression for the timing behavior of a locked oscillator with tuning error (subject of course to the usual small angle restrictions) for any arbitrary pulse pattern:

$$\vec{\epsilon}^{\,o}(n') - \vec{\epsilon}^{\,o}(n' - 1) = \frac{1}{A}\left[\vec{\epsilon}^{\,i}(n') - \vec{\epsilon}^{\,o}(n' - 1) + \frac{\delta f}{F} b_{n'}\right] - \frac{\delta f}{F} b_{n'}. \quad (37)$$

The parameter A is now determined by the ratio of the amplitudes of the oscillator wave and the synchronizing signal. Equation (37) shows that

a locked oscillator timing circuit with tuning error may be treated rigorously as a linear transducer to the timing deviation for any arbitrary pulse pattern; the equivalent linear transducer has two inputs, $\bar{\epsilon}^i(n')$ and $b_{n'}$.

Comparison with the results of the preceding section shows that for zero tuning error (37) also describes the approximate timing response of a tuned circuit timing filter, where A is now the average timing wave amplitude, as in (36). Unfortunately, (37) does *not* provide a valid approximation to the behavior of a tuned circuit when tuning error is present. This is best shown by comparing a tuned circuit and the corresponding locked oscillator in a simple case.

Consider a random pulse pattern where the probability that any time slot contains a pulse is very close to one, independently for each time slot. Only rarely will vacant time slots lie close together in this pulse pattern, and so it suffices to study the effect of a single missing pulse, as shown in Fig. 6(a). For $n' \leqq 0$ both the output timing deviation and the timing wave amplitude are constant for both the locked oscillator and the tuned circuit. Making A for the locked oscillator equal to the

$$b_{n'} = \begin{cases} 2, & n'=1 \\ 1, & n' \neq 1 \end{cases}$$

(a)

(b)
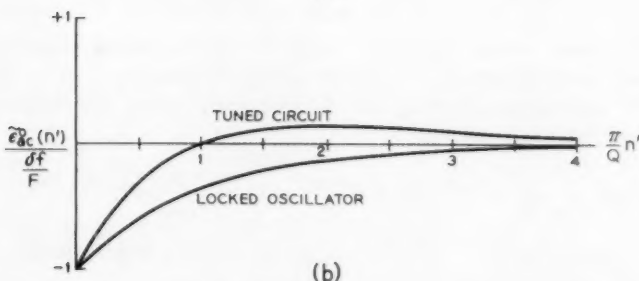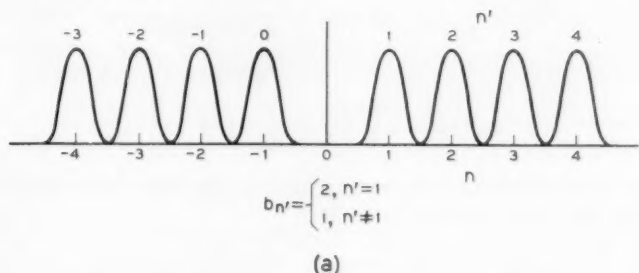
Fig. 6 — Output timing deviation for locked oscillator and tuned-circuit timing filters with tuning error, for single missing pulse: (a) pulse pattern; (b) output timing deviations.

steady-state timing wave amplitude for the tuned circuit, we have for both cases, from (24) and (35) to (37):

$$\tilde{A}_{n'} = A = \sum_{k=0}^{\infty} e^{-(\pi/Q)k} = \frac{1}{1 - e^{-\pi/Q}}; \qquad n' \le 0, \tag{38}$$

$$\bar{\epsilon}^{\circ}(n') = \bar{\epsilon}_{dc}^{\circ} = -\frac{\delta f}{F} A \left(1 - \frac{1}{A}\right) = -\frac{\delta f}{F} \frac{e^{-\pi/Q}}{1 - e^{-\pi/Q}}; \qquad n' \le 0. \tag{39}$$

The transient or ac component of the output timing deviation, $\bar{\epsilon}_{ac}^{\circ}(n')$, is of principal interest, where

$$\bar{\epsilon}^{\circ}(n') = \bar{\epsilon}_{dc}^{\circ} + \bar{\epsilon}_{ac}^{\circ}(n'). \tag{40}$$

Then, from (35), we have

$$\bar{\epsilon}_{ac}^{\circ}(n') - \bar{\epsilon}_{ac}^{\circ}(n' - 1)$$

$$= -\frac{1}{\tilde{A}_{n'}} \bar{\epsilon}_{ac}^{\circ}(n' - 1) + \frac{\delta f}{F} \frac{A - \tilde{A}_{n'}}{\tilde{A}_{n'}}; \qquad n' > 1, \tag{41}$$

$$\tilde{A}_{n'} = \begin{cases} A & ; & \text{locked oscillator} \\ A - e^{-(\pi/Q)n'}; & \text{tuned circuit.} \end{cases}$$

The initial condition for the difference equation (41) is given by

$$\bar{\epsilon}_{ac}^{\circ}(1) = -\frac{\delta f}{F} \left[2 - \frac{1}{\tilde{A}_1}(A + 1)\right]. \tag{42}$$

For the locked oscillator, the second term on the right-hand side of (41) vanishes and, subject to (42), we find

$$\bar{\epsilon}_{ac}^{\circ}(n') = -\frac{\delta f}{F} e^{-(\pi/Q)n'}; \qquad \text{locked oscillator.} \tag{43}$$

For the tuned circuit, the second term of (41) no longer vanishes, but adds an inhomogeneous term to the solution. From (41), $A - \tilde{A}_{n'} = e^{-(\pi/Q)n'}$ and, approximating the $\tilde{A}_{n'}$ occurring in the denominators of (41) and (42) by A, we find

$$\bar{\epsilon}_{ac}^{\circ}(n') = -\frac{\delta f}{F} e^{-(\pi/Q)n'}[(2 - e^{-\pi/Q}) - (1 - e^{-\pi/Q})n']; \tag{44}$$

$$\text{tuned circuit.}$$

For the high-Q case, this may be further simplified:

$$\bar{\epsilon}_{ac}^{\circ}(n') = -\frac{\delta f}{F} e^{-(\pi/Q)n'} \left[1 - \frac{\pi}{Q} n'\right]; \qquad \text{tuned circuit, } Q \gg 1. \tag{45}$$

These results may alternately be obtained by inspection from the vector diagram of Fig. 5.

Equations (43) and (45) are plotted in Fig. 6(b). For the tuned circuit, $\bar{\epsilon}_{ac}^{\,o}(n')$ initially approaches zero twice as fast as for the locked oscillator. Consequently, a tuned circuit and the corresponding locked oscillator are approximately equivalent only for zero tuning error; when tuning error is present their behavior is no longer similar. In the present paper we are able to treat the effect of tuning error in a chain of repeaters only for locked-oscillator timing circuits.

2.4 Average Value and Fluctuation of the Timing Wave Amplitude

We now determine the average value and standard deviation of the timing wave amplitude $\tilde{A}_{n'}$, denoted by $A = \langle \tilde{A}_{n'} \rangle$ and σ_A respectively, for a tuned circuit. The average value is needed in (36), which gives the approximate timing response for a tuned circuit timing filter with zero tuning error. The standard deviation indicates the magnitude of the departure of $\tilde{A}_{n'}$ from its average value A, which is neglected in this approximate analysis.

Consider a random pulse pattern with the probability p that any time slot contains a pulse, independently for each time slot. Equation (24) gives A_n; for the a_{n-k}'s we have

$$a_n = 1,$$

$$\left.\begin{array}{l} \Pr[a_{n-k} = 1] = p \\ \Pr[a_{n-k} = 0] = 1 - p \end{array}\right\} k > 1. \tag{46}$$

The different a's are independent. Then, from (24), we have

$$A = \langle \tilde{A}_{n'} \rangle = \sum_{k=0}^{\infty} \langle a_{n-k} \rangle e^{-(\pi/Q)k} = 1 + \frac{pe^{-\pi/Q}}{1 - e^{-\pi/Q}}, \tag{47}$$

$$\begin{aligned} \langle \tilde{A}_{n'}^{\,2} \rangle &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \langle a_{n-k} a_{n-l} \rangle e^{-(\pi/Q)(k+l)} \\ &= 1 + 2p \frac{e^{-\pi/Q}}{1 - e^{-\pi/Q}} + \left(\frac{pe^{-\pi/Q}}{1 - e^{-\pi/Q}} \right)^2 + p(1 - p) \frac{e^{-2\pi/Q}}{1 - e^{-2\pi/Q}}, \end{aligned} \tag{48}$$

$$\sigma_A^{\,2} = \langle \tilde{A}_{n'}^{\,2} \rangle - \langle \tilde{A}_{n'} \rangle^2 = p(1 - p) \frac{e^{-2\pi/Q}}{1 - e^{-2\pi/Q}}. \tag{49}$$

For the high-Q case

$$A = \frac{pQ}{\pi}, \quad \sigma_A^2 = p(1 - p)\frac{Q}{2\pi}, \quad \left(\frac{\sigma_A}{A}\right)^2 = \frac{1 - p}{p}\frac{\pi}{2Q}; \quad Q \gg 1. \quad (50)$$

This agrees with the results of a similar calculation made by W. R. Bennett.[7]

The timing-wave amplitude will be close to its average value with high probability if

$$\frac{\sigma_A}{A} \ll 1, \quad Q \gg \frac{1 - p}{p}. \quad (51)$$

The smaller the probability that a time slot contains a pulse, the larger Q must be in order for the variation in the timing wave amplitude to be small.

For any periodic pulse pattern, $\tilde{A}_{n'}$ is easily determined as a periodic function of n' from (24). For a high enough Q the average timing wave amplitude A is again given by (50), where p is now the average number of time slots containing pulses. In the special case of a periodic pattern with every Mth pulse present, $\tilde{A}_{n'}$ is of course constant, given by (31) or (33).

2.5 *Tuned Circuit Excited by Raised Cosine Pulses*

Up to now we have considered only the case of a timing filter excited by impulses. However, in the simplest repeaters the signal pulses themselves or their rectified envelopes (in baseband or carrier systems, respectively) would be used to drive the timing circuit. The finite width of these driving pulses may add an additional component of timing noise at the output of each repeater. We now consider the behavior of a tuned circuit excited by raised cosine pulses, which approximate the pulse shape that might be used in a practical system.

A driving pulse centered at $t = 0$ is given by

$$p(sFt) = \frac{1}{2}(1 + \cos 2\pi sFt); \quad |Ft| < \frac{1}{2s}, \quad (52)$$

where F is the pulse repetition frequency and s is a parameter determining the pulse width. For $s = 1$ adjacent pulses are just resolved; for $s = \frac{2}{3}$ adjacent pulses overlap such that the amplitude half way between sample points is equal to one-half the peak pulse amplitude. These two cases are illustrated in Fig. 7. The complex response $P(t)$ of a tuned cir-

$$p(Ft) = 1/2 \left[1 + \cos 2\pi Ft \right] \qquad p(2/3\ Ft) = 1/2 \left[1 + \cos 2/3\ 2\pi Ft \right]$$
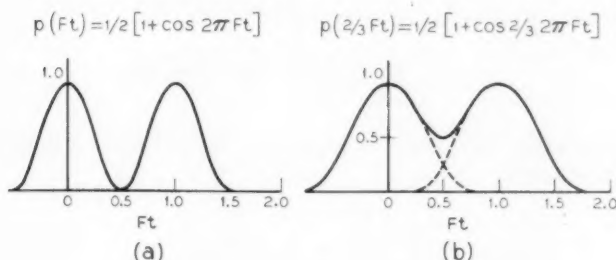
Fig. 7 — Driving pulses for timing circuit: (a) pulses resolved; (b) pulses over-lapping.

cuit to this pulse is

$$P(t) = \int_{-\infty}^{t} p(sF\tau)H(t - \tau)\, d\tau, \tag{53}$$

where $H(t)$, given in (1) and (4), is the complex impulse response of the tuned circuit. Thus,

$$P(t) = \frac{1}{2F} I_s(Ft)H(t), \tag{54}$$

where

$$I_s(Ft) = \begin{cases} 0; & Ft < -\dfrac{1}{2s} \\[2ex] \displaystyle\int_{-1/2s}^{Ft} (1 + \cos 2\pi s\tau)e^{+(\pi/Q)(1+\delta f/F)\tau}e^{-j2\pi(1+\delta f/F)\tau}\, d\tau; & \\[1ex] & -\dfrac{1}{2s} < Ft < +\dfrac{1}{2s} \\[2ex] I_s\left(\dfrac{1}{2s}\right); & +\dfrac{1}{2s} < Ft. \end{cases} \tag{55}$$

The pulse train applied to the tuned circuit now becomes [from (5) through (9)]:

$$g(t) = \sum_{n=-\infty}^{\infty} a_n p\{s[Ft - n - \epsilon^i(n)]\}. \tag{56}$$

The tuned-circuit response $G(t)$ to all pulses up to and including the

pulse in the mth time slot is given by

$$G(t) = \sum_{n=-\infty}^{m} a_n P[t - nT - \epsilon^i(n)T], \qquad (57)$$

corresponding to (10), and $G(t)$ may again be written in a form similar to (12) and (13):

$$G(t) = \frac{I_s\left(\frac{1}{2s}\right)}{2FC}\left(1 + \frac{j}{2Q}\right) A(t) e^{j2\pi Ft}, \qquad (58)$$

$$A(t) = a(t) e^{j\varphi(t)}, \qquad$$

where, as before, $|A(t)| = a(t)$ is the normalized amplitude and $\angle A(t) = \varphi(t)$ is the phase deviation of the timing wave. Corresponding to (13), we now have for $A(t)$

$$A(t) = \sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(1+\delta f/F)[Ft-n-\epsilon^i(n)]} \cdot e^{j2\pi\{(1+\delta f/F)[Ft-n-\epsilon^i(n)]-Ft\}}$$

$$+ a_m \left[\frac{I_s[Ft - m - \epsilon^i(m)]}{I_s\left(\frac{1}{2s}\right)} - 1\right] e^{-(\pi/Q)(1+\delta f/F)[Ft-m-\epsilon^i(m)]} \qquad (59)$$

$$\cdot e^{j2\pi\{(1+\delta f/F)[Ft-m-\epsilon^i(m)]-Ft\}};$$

$$(m - 1) + \epsilon^i(m - 1) + \frac{1}{2s} < Ft < (m + 1) + \epsilon^i(m + 1) - \frac{1}{2s},$$

where t is restricted such that only the pulse in the mth time slot differs from 0.

We now assume that the phase of the timing wave $\varphi(t)$ at the negative-going zero crossing following the peak of each received pulse determines the timing deviation of the corresponding timing pulse. In the analysis for impulse excitation in Section 2.1 the timing wave phase was evaluated just after the reception of a driving impulse, since a different choice would introduce only a small dc error in the output timing deviation. In contrast, with driving pulses of finite width the choice of time at which $\varphi(t)$ is evaluated has an appreciable effect on the last term of (59), which gives the additional output timing deviation due to finite pulse width. Thus, evaluating (59) at

$$Ft = F\left(t_m + \frac{T}{4}\right) = m + \epsilon^i(m) + \frac{1}{4}, \qquad (60)$$

where t_n is given in (9), we obtain

$$A\left(t_m + \frac{T}{4}\right) = [e^{-(\pi/Q)(1+\delta f/F)1/4}e^{j(\delta f/F)\pi/2}]$$

$$\cdot e^{-j2\pi\epsilon^i(m)}\left\{\sum_{n=-\infty}^{m} a_n e^{-(\pi/Q)(1+\delta f/F)[(m-n)+\epsilon^i(m)-\epsilon^i(n)]}\right.$$

(61)

$$\left.\cdot e^{j2\pi\{\epsilon^i(m)-\epsilon^i(n)+(\delta f/F)[(m-n)+\epsilon^i(m)-\epsilon^i(n)]\}} + a_m\left[\frac{I_s\left(\frac{1}{4}\right)}{I_s\left(\frac{1}{2s}\right)} - 1\right]\right\},$$

corresponding to (16). In order for the restriction of (59) to be satisfied s must be greater than $\frac{2}{3}$. The terms in the first bracket arise from the fact that we have determined the phase one-quarter of a pulse period after the peak of the received pulse, rather than just after the peak as in Section 2.1; they represent a small decrease in amplitude and a small dc phase shift, and may be neglected. Then, making use of the discussion in Section 2.1 and assuming as before that the alignment deviation $\alpha(n) = \epsilon^o(n) - \epsilon^i(n)$ remains small, we obtain from (17) and (61) the normalized timing wave amplitude and output timing deviation (with complete retiming):

$$A_n = \sum_{k=0}^{\infty} a_{n-k}e^{-(\pi/Q)k},$$

(62)

$$\epsilon^o(n) = \frac{\sum_{k=0}^{\infty} a_{n-k}e^{-(\pi/Q)k}\left[\epsilon^i(n-k) - \frac{\delta f}{F}k\right]}{A_n} - w_s\frac{a_n}{A_n},$$

(63)

where w_s depends on the pulse width and is given by

$$w_s = \frac{1}{2\pi}\operatorname{Im}\left[\frac{I_s\left(\frac{1}{4}\right)}{I_s\left(\frac{1}{2s}\right)} - 1\right].$$

(64)

These results are again subject to the restrictions of (20) and (21). The following approximate results for the high-Q case may be obtained from

(55):

$$s = 1,$$

$$w_1 = 0.0253 + \frac{0.0325}{Q},$$

$$I_1\left(\frac{1}{4}\right) = 0.534 + \frac{0.185}{Q} + j\left[0.0795 - \frac{0.298}{Q}\right],$$

$$I_1\left(\frac{1}{2}\right) = 0.5\left[1 - j\frac{0.75}{Q}\right].$$

$$s = \frac{2}{3},$$

$$w_{2/3} = 0.1038 + \frac{0.349}{Q},$$

$$I_{2/3}\left(\frac{1}{4}\right) = 0.430 + \frac{0.235}{Q} + j\left[0.166 - \frac{0.429}{Q}\right],$$

$$I_{2/3}\left(\frac{3}{4}\right) = 0.255\left[1 - j\frac{2.3}{Q}\right].$$

Except for the last term, (63) is identical to (25) for impulse excitation; the last term represents the additional output timing deviation caused by the finite pulse width. This timing deviation is inversely proportional to the instantaneous timing wave amplitude, and is thus quite similar to the effect of amplitude-to-phase conversion in the limiter of the timing system.[2, 8] In both cases an identical timing deviation will be added at the output of every repeater of the chain, since the pulse pattern is the same for every repeater. This additional timing deviation is caused by the fact that the driving pulse is not zero when the timing wave goes through zero. However, the zero crossings corresponding to vacant time slots are not disturbed; W. M. Goodall has pointed out that inverting the pulse pattern in each repeater will therefore eliminate this source of timing noise. If the driving pulse is short enough so that it falls to zero before the zero crossing of the timing wave occurs, no additional timing deviation will be produced; in particular, the driving pulses may be impulses as above or square pulses of length $T/2$.[7] Finally, a weakly coupled locked oscillator will not be affected by the finite width of the synchronizing pulses.

III. FOURIER ANALYSIS OF DISCRETE FUNCTIONS

3.1 Transforms of Discrete Functions

The timing deviation $\epsilon(n)$ or $\bar{\epsilon}(n')$ and the various other quantities discussed above may be called discrete real functions, since they are defined only for integral values of their argument. The Fourier analysis of discrete functions is quite similar to that for continuous functions and is summarized below.

The transform $X(f)$ of a discrete function $x(n)$ is defined by[9, 10]

$$X(f) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi fn}.$$ (65)

The inverse transform is then

$$x(n) = \int_{-1/2}^{+1/2} X(f)e^{+j2\pi fn}\, df.$$ (66)

This transformation is closely related to the z-transform[11, 12] or generating function. Equations (65) and (66) may be regarded as a statement of the Fourier series theorem, in which the usual roles of function and transform have been reversed.

The following table [Equation (67)] summarizes some of the properties of discrete transforms that will be of use; $x_1(n)$ and $x_2(n)$ are any two different discrete functions with corresponding transforms $X_1(f)$ and $X_2(f)$:

Function	Transform	
$x_1(n)$	$X_1(f)$	(67a)
$x_2(n)$	$X_2(f)$	(67b)
$x_1(n + k)$	$e^{j2\pi fk}X_1(f)$	(67c)
$\Delta x_1(n) = x_1(n + 1) - x_1(n)$	$(e^{j2\pi f} - 1)\, X_1(f)$	
	$\quad = 2je^{j\pi f} \sin \pi f X_1(f)$	(67d)
$x_1(n)\otimes x_2(n) = \displaystyle\sum_{k=-\infty}^{\infty} x_1(n - k)x_2(k)$	$X_1(f)X_2(f)$	
$\quad = \displaystyle\sum_{k=-\infty}^{\infty} x_1(k)x_2(n - k)$		(67e)
$x_1(n)x_2(n)$	$X_1(f)\otimes X_2(f)$	
	$\quad = \displaystyle\int_{-1/2}^{+1/2} X_1(f - \tau)X_2(\tau)\, d\tau$	
	$\quad = \displaystyle\int_{-1/2}^{+1/2} X_1(\tau)X_2(f - \tau)\, d\tau$	(67f)
$\varphi_1(k) = \displaystyle\sum_{n=-\infty}^{\infty} x_1(n)x_1(n + k)$	$E_1(f) = X_1(f)X_1{}^*(f)$	
	$\quad = \lvert X_1(f)\rvert^2.$†	(67g)

---

† The * denotes the complex conjugate.

In (67g) the autocorrelation function $\varphi(k)$ and the energy spectrum $E(f)$ of a discrete function $x(n)$ have been defined; the fact that they are transforms of each other follows from the convolution theorem (67e) and the fact that $x(n)$ is real. The total energy E of a discrete function $x(n)$ is defined by

$$E = \sum_{n=-\infty}^{\infty} x^2(n). \tag{68}$$

Then, from (67g),

$$E = \varphi(0) = \int_{-1/2}^{1/2} E(f)\, df = \int_{-1/2}^{1/2} |X(f)|^2\, df. \tag{69}$$

The results of the present section apply only to discrete functions having a finite energy E as defined in (68), so that their transforms exist. Discrete random functions, which have infinite energy but a finite average power, will be discussed in Section 3.3.

3.2 *Discrete Transducers*

A discrete linear invariant transducer may be characterized by its impulse response. A discrete unit impulse at $n = n_0$ is defined by

$$\delta(n - n_0) = \begin{cases} 1; & n = n_0 \\ 0; & n \neq n_0, \end{cases} \tag{70}$$

corresponding to the delta function in the continuous case. Let $h(n)$ be the output of a discrete linear invariant transducer for a unit impulse input at the origin; for stable systems

$$h(n) = 0; \quad n < 0. \tag{71}$$

Then the output $x^{\circ}(n)$ for an arbitrary input $x^{i}(n)$ is given by the convolution of the input and the impulse response $h(n)$ of the transducer:

$$x^{\circ}(n) = \sum_{k=-\infty}^{n} x^{i}(k)h(n - k)$$

$$= \sum_{k=0}^{\infty} x^{i}(n - k)h(k). \tag{72}$$

Let the transforms according to (65) and (66) of $h(n)$, $x^{i}(n)$ and $x^{\circ}(n)$ be $H(f)$, $X^{i}(f)$ and $X^{\circ}(f)$ respectively; $H(f)$ is the frequency response or transfer function of the transducer. Then, from (67e), we have

$$X^{\circ}(f) = H(f)X^{i}(f). \tag{73}$$

If the input consists of a single exponential of frequency f,

$$x^i(n) = A_i e^{j2\pi f n}, \tag{74}$$

then the output is an exponential of the same frequency,

$$x^o(n) = A_o e^{j2\pi f n}, \tag{75}$$
$$A_o = H(f)A_i,$$

as in the continuous case.

Finally, consider two cascaded discrete transducers with individual impulse responses $h_1(n)$ and $h_2(n)$ and corresponding transfer functions $H_1(f)$ and $H_2(f)$. The over-all impulse response $h(n)$ is equal to the convolution of $h_1(n)$ and $h_2(n)$, and the over-all transfer function $H(f)$ is equal to the product of $H_1(f)$ and $H_2(f)$, as in the continuous case:

$$h(n) = \sum_{k=0}^{n} h_1(n-k)h_2(k), \tag{76}$$

$$H(f) = H_1(f)H_2(f). \tag{77}$$

3.3 *Discrete Random Functions*

A discrete stationary random function $x(n)$ will have infinite energy but a finite average power P given by

$$P = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x^2(n). \tag{78}$$

The covariance $\rho(k)$ is defined as

$$\rho(k) = \langle x(n)x(n+k) \rangle, \tag{79}$$

where $\langle \rangle$ denotes an average over the ensemble. For a stationary ergodic ensemble the power spectrum $P(f)$ and the covariance $\rho(k)$ are transforms of each other, according to (65) and (66):

$$P(f) = \sum_{k=-\infty}^{\infty} \rho(k)e^{-j2\pi f k}, \tag{80}$$

$$\rho(k) = \int_{-1/2}^{1/2} P(f)e^{+j2\pi f k}\, df. \tag{81}$$

The average power P is the integral of the power spectrum

$$P = \int_{-1/2}^{1/2} P(f)\, df. \tag{82}$$

A white noise of power N_0 has a covariance

$$\rho(k) = N_0 \delta(k) = \begin{cases} N_0; & k = 0 \\ 0; & k \neq 0. \end{cases} \tag{83}$$

Thus the power spectrum is

$$P(f) = N_0. \tag{84}$$

This process is called a white noise because the power spectrum is constant with frequency. The values of this noise for different values of the independent variable n are uncorrelated; this random process is used to describe the equivalent timing noise added at the input of each repeater, as discussed in Section I.

IV. TIMING IN A CHAIN OF REPEATERS WITH ZERO TUNING ERROR AND COMPLETE RETIMING

4.1 *Introduction*

In this section we determine the timing, spacing and alignment noise power spectra and total powers (mean square values) caused by the input noise at each repeater for a chain of repeaters with zero tuning error and complete retiming. The repeaters may employ either tuned-circuit or locked-oscillator timing circuits.

From (36) or (37) the response of each repeater is given by

$$\bar{\epsilon}^o(n') - \bar{\epsilon}^o(n' - 1) = \frac{1}{A} [\bar{\epsilon}^i(n') - \bar{\epsilon}^o(n' - 1)],$$

$$A = \langle \tilde{A}_{n'} \rangle. \tag{85}$$

For locked oscillators, (85) is exact. The normalized timing wave amplitude A is equal to the ratio of the amplitude of the oscillator wave to the amplitude of the transient started by a single signal pulse.

For tuned circuits we assume that the driving pulses for the timing circuit are short enough so that no additional output timing noise results from the finite pulse width, as discussed in Section 2.5, and we study only the effects of input noise. Equation (85) will be exact only for periodic pulse patterns with every Mth pulse present (including all pulses present); for random or general periodic pulse patterns it gives a good approximation for the response of a repeater. The normalized timing wave amplitude $\tilde{A}_{n'}$ is given by (24); A is given by (31) or (50).

Let the discrete transforms of the input and output timing deviations

$\tilde{\epsilon}^i(n')$ and $\tilde{\epsilon}^o(n')$ be $\tilde{C}^i(f)$ and $\tilde{C}^o(f)$ respectively, where the symbol ~ indicates that the transforms have been taken with respect to the independent variable n' [Fig. 4(a)]. Then, transforming (85) by the use of (67), we obtain for the transfer function $\tilde{H}(f)$ of a single repeater

$$\tilde{H}(f) = \frac{\tilde{C}^o(f)}{\tilde{C}^i(f)} = \frac{1/A}{1 - (1 - 1/A)e^{-j2\pi f}}. \tag{86}$$

It is convenient to define $\tilde{Q}$ as the effective "Q" of the timing circuit to the timing deviations, as a function of the average timing wave amplitude A:

$$A = \frac{1}{1 - e^{-\pi/\tilde{Q}}}. \tag{87}$$

In the high-$\tilde{Q}$ case,

$$\tilde{Q} = \pi A, \qquad \tilde{Q} \gg 1. \tag{88}$$

For tuned-circuit timing filters $\tilde{Q}$ may be conveniently expressed in terms of the tuned circuit Q. For a periodic pulse pattern with every Mth pulse present, from (31),

$$\tilde{Q} = \frac{Q}{M}. \tag{89}$$

For all pulses present, $M = 1$ and $\tilde{Q} = Q$. For random or general periodic pulse patterns (50) yields, for the high-Q case,

$$\tilde{Q} = pQ, \qquad \tilde{Q} \gg 1, \tag{90}$$

where p is either the probability that a time slot contains a pulse or the average number of time slots containing pulses.

In terms of $\tilde{Q}$, the frequency response of a repeater becomes

$$\tilde{H}(f) = \frac{1 - e^{-\pi/\tilde{Q}}}{1 - e^{-\pi/\tilde{Q}}e^{-j2\pi f}}. \tag{91}$$

Taking the inverse transform of $\tilde{H}(f)$, the impulse response $\tilde{h}(n')$ of a single repeater is

$$\tilde{h}(n') = (1 - e^{-\pi/\tilde{Q}})e^{-(\pi/\tilde{Q})n'}; \qquad n' \geq 0. \tag{92}$$

Finally, the square of the absolute magnitude of the frequency response of a single repeater may be written

$$| \tilde{H}(f) |^2 = \frac{1}{1 + \operatorname{csch}^2 \dfrac{\pi}{2\tilde{Q}} \sin^2 \pi f}. \tag{93}$$

Equation (93) is plotted in Fig. 8(a) for the high-$\tilde{Q}$ case. The low-frequency components of the timing deviation are transmitted with little loss in amplitude, and the high-frequency components are substantially suppressed in a single repeater. The 3-db bandwidth is given by

$$f_{3db} = \frac{1}{2\tilde{Q}}; \qquad \tilde{Q} \gg 1. \tag{94}$$

4.2 Impulse and Frequency Response of a Chain of Repeaters

The transfer function of K repeaters in cascade, $\tilde{H}_K(f)$, is equal to the Kth power of the transfer function of a single repeater. From (77)

(a)

(b)
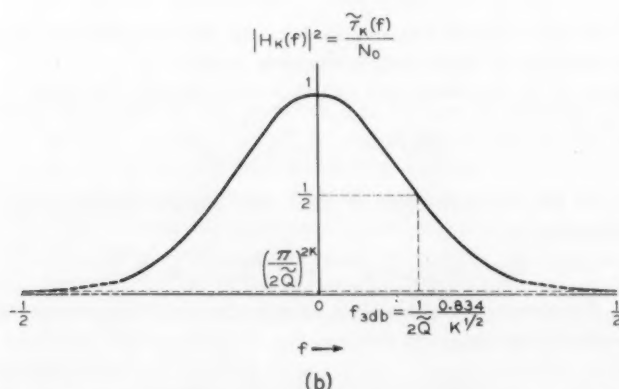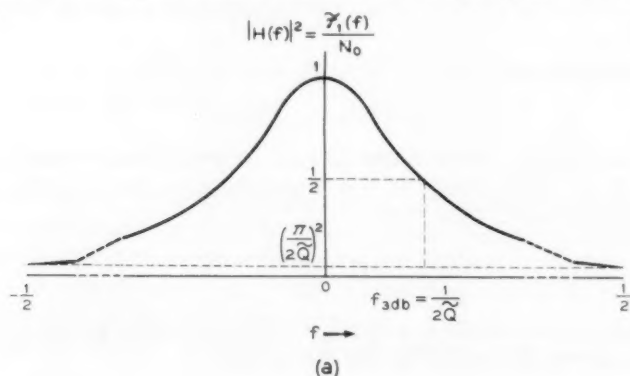
Fig. 8 — Transfer function for repeaters with zero tuning error and complete retiming, with $\tilde{Q} \gg 1$: (a) single repeater; (b) k cascaded repeaters, $k \gg 1$.

and (91),

$$\tilde{H}_K(f) = \tilde{H}^K(f) = \left[\frac{1 - e^{-\pi/\tilde{Q}}}{1 - e^{-\pi/\tilde{Q}}e^{-j2\pi f}} \right]^K. \tag{95}$$

From (93),

$$|\tilde{H}_K(f)|^2 = |\tilde{H}(f)|^{2K} = \left[\frac{1}{1 + \operatorname{csch}^2 \dfrac{\pi}{2\tilde{Q}} \sin^2 \pi f} \right]^K. \tag{96}$$

Equation (96) is plotted in Fig. 8(b); the 3-db bandwidth for the high-$\tilde{Q}$ case is

$$f_{3db} = \frac{1}{2\tilde{Q}} \sqrt{2^{1/K} - 1}; \qquad \tilde{Q} \gg 1. \tag{97}$$

For a long chain of repeaters (97) becomes

$$f_{3db} = \frac{1}{2\tilde{Q}} \frac{0.834}{K^{1/2}}; \qquad \tilde{Q} \gg 1, \qquad K \gg 1. \tag{98}$$

The bandwidth of a chain of repeaters to the timing deviation varies inversely as the square root of the number of repeaters.

The impulse response for K cascaded repeaters, $\tilde{h}_K(n')$, may be found either by taking the inverse transform of the frequency response, given in (95), or by finding the $(K - 1)$-fold convolution of the impulse response for a single repeater, given in (92). Using the latter method,

$$\tilde{h}_K(n') = (1 - e^{-\pi/\tilde{Q}})^K e^{-(\pi/\tilde{Q})n'} \underbrace{\sum_{n_{K-1}=0}^{n'} \sum_{n_{K-2}=0}^{n_{K-1}} \cdots \sum_{n_2=0}^{n_3} \sum_{n_1=0}^{n_2} 1.}_{K - 1 \text{ summations}} \tag{99}$$

This expression is easily evaluated using functions occurring in the theory of difference equations.[13] The function $k^{[m]}$ is defined as follows:

$$k^{[m]} = k(k - 1)(k - 2) \cdots (k - m + 1) = \frac{\Gamma(k + 1)}{\Gamma(k - m + 1)}. \tag{100}$$

The brackets around the superscript m indicate that it is not to be regarded as an exponent. Note that

$$k^{[0]} = 1. \tag{101}$$

This function behaves with respect to difference and summation operators in a similar manner to the power function x^m with respect to differential and integral operators. In particular,

$$\Delta k^{[m]} = (k + 1)^{[m]} - k^{[m]} = mk^{[m-1]},$$

$$\sum_{k=M}^{N} k^{[m]} = \left[\frac{k^{[m+1]}}{m + 1} \right]_{M}^{N+1}. \tag{102}$$

Substituting these relations into (99), we obtain

$$\tilde{h}_K(n') = (1 - e^{-\pi/\tilde{Q}})^K e^{-(\pi/\tilde{Q})n'} \frac{\Gamma(n' + K)}{\Gamma(K)\Gamma(n' + 1)} \tag{103}$$

for the impulse response of K repeaters in cascade. For $K = 1$ this expression becomes identical to (92). For a single repeater the impulse response decreases exponentially to zero; for a chain of repeaters the impulse response starts from zero, increases to a maximum value, and decreases again to zero.

4.3 *Timing, Spacing and Alignment Noise*

Consider a chain of N repeaters with an independent white timing noise of average power N_0 [(83) and (84)] added at the input of every repeater. The output timing noise power spectrum is equal to the sum of the power spectra produced by the individual noise sources, since the system is linear and the various noise sources are independent.

Let $\tilde{\tau}_K(f)$ be the power spectrum of the timing noise at the output of a chain of K repeaters with a white timing noise of power N_0 introduced at the input of only the first repeater. Then, from (73) and (96),

$$\tilde{\tau}_K(f) = N_0 \mid \tilde{H}(f) \mid^{2K}. \tag{104}$$

This noise spectrum has the same shape as the transfer function of (93) or (96), shown on Fig. 8, so these curves also show $\tilde{\tau}_K(f)$. The timing noise power spectrum $\tilde{T}_N(f)$ at the output of a chain of N repeaters with an independent white timing noise of power N_0 added at the input of each repeater is now

$$\tilde{T}_N(f) = \sum_{K=1}^{N} \tilde{\tau}_K(f) = N_0 \frac{1 - \mid \tilde{H}(f) \mid^{2N}}{\mid \tilde{H}(f) \mid^{-2} - 1}, \tag{105}$$

where $\mid \tilde{H}(f) \mid^2$ is given in (93).

Fig. 9(a) shows $\tilde{T}_N(f)$ for a high $\tilde{Q}$ and a fairly long chain of repeaters. At zero frequency

$$\tilde{T}_N(0) = N_0 N. \tag{106}$$

The low-frequency noise components from each noise source are unattenuated, and their powers add at the output of the repeater chain. The

high-frequency components are strongly attenuated in each repeater. The 3-db bandwidth in the high-$\tilde{Q}$ case is

$$f_{3db} = \frac{1}{2\tilde{Q}} \frac{1.265}{N^{1/2}}; \qquad \tilde{Q} \gg 1, \qquad N \gg 1. \tag{107}$$

As the length of the repeater chain approaches infinity, the timing-noise power spectrum approaches a limiting form

$$\lim_{N \to \infty} \tilde{T}_N(f) = N_0 \frac{\sinh^2 \frac{\pi}{2\tilde{Q}}}{\sin^2 \pi f}, \tag{108}$$

as indicated in Fig. 9(a). In this limiting case, for each narrow band the decrease in noise power in passing through each repeater is equal to the noise power added at the input of the next repeater.

The total timing noise power is given simply by the integral of the power spectrum, according to (82). Thus,

$$\bar{\tau}_K = \int_{-1/2}^{1/2} \tilde{\tau}_K(f) \, df,$$
$$\tilde{T}_N = \int_{-1/2}^{1/2} \tilde{T}_N(f) \, df, \tag{109}$$

where $\bar{\tau}_K$ is the total timing noise power at the output of a chain of repeaters with noise added at the input of only the first repeater, $\tilde{T}_N$ is the total timing noise power at the output of a chain of repeaters with noise added at the input of every repeater and the power spectra $\tilde{\tau}_K(f)$ and $\tilde{T}_N(f)$ are given by (104) and (105). While exact expressions for $\bar{\tau}_K$ and $\tilde{T}_N$ can be found, they are so complicated as to be of little use except for short repeater chains, and are not suitable for considering the high-$\tilde{Q}$ case. An asymptotic expression for $\tilde{T}_N$ has been found by S. O. Rice and is derived in the Appendix; a similar analysis has been performed for $\bar{\tau}_N$, but only the final result is given here. For the high-$\tilde{Q}$ case:

$$\bar{\tau}_K = N_0 \frac{1}{\pi} \sinh \frac{\pi}{2\tilde{Q}} \frac{\Gamma(\frac{1}{2}) \, \Gamma(K - \frac{1}{2})}{\Gamma(K)}; \qquad K \geqq 2, \qquad \tilde{Q} \gg 1, \tag{110}$$

$$\tilde{T}_N = N_0 \frac{2N}{\pi} \sinh \frac{\pi}{2\tilde{Q}} \frac{\Gamma(\frac{1}{2}) \, \Gamma(N + \frac{1}{2})}{\Gamma(N + 1)}; \qquad \tilde{Q} \gg 1. \tag{111}$$

For a long chain of repeaters Stirling's approximation yields:

$$\bar{\tau}_K = N_0 \frac{1}{2\tilde{Q}} \sqrt{\frac{\pi}{K - 1}}; \qquad \tilde{Q} \gg 1, \qquad K \gg 1, \tag{112}$$
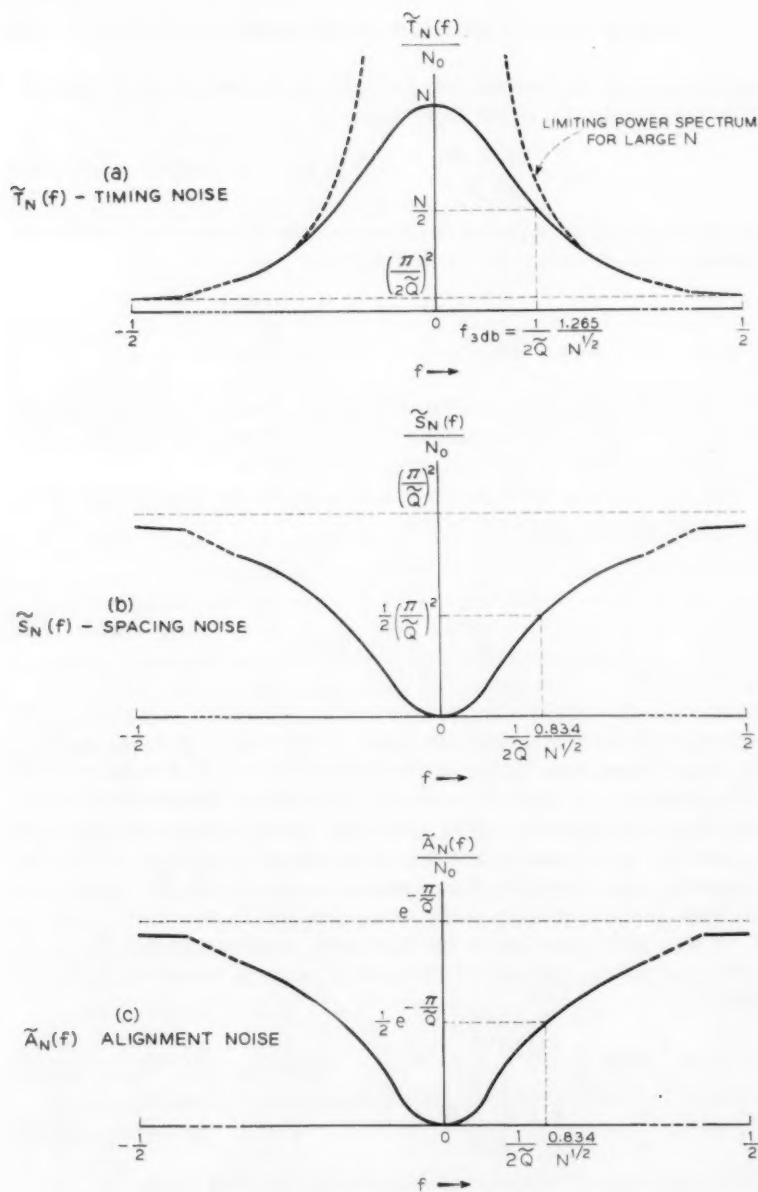
Fig. 9 — Timing (a), spacing (b) and alignment (c) noise power spectra for a chain of N repeaters with zero tuning error and complete retiming, with a white timing noise of power N_0 added at the input of each repeater; $Q \gg 1$, $N \gg 1$.

$$\tilde{T}_N = N_0 \frac{\sqrt{\pi N}}{\tilde{Q}}; \qquad\qquad \tilde{Q} \gg 1, \qquad N \gg 1. \tag{113}$$

Finally, for a short repeater chain the exact results are reasonably simple:

$$\tilde{T}_1 = \bar{\tau}_1 = N_0 \tanh \frac{\pi}{2\tilde{Q}},$$

$$\bar{\tau}_2 = N_0 \tanh^2 \frac{\pi}{2\tilde{Q}} \operatorname{ctnh} \frac{\pi}{\tilde{Q}}. \tag{114}$$

For the high-$\tilde{Q}$ case

$$\tilde{T}_1 = \bar{\tau}_1 = N_0 \frac{\pi}{2\tilde{Q}}; \qquad \tilde{Q} \gg 1,$$

$$\bar{\tau}_2 = N_0 \frac{\pi}{4\tilde{Q}}; \qquad \tilde{Q} \gg 1. \tag{115}$$

The spacing and alignment noise are easily found in terms of the above results. Let $\tilde{S}_N(f)$ and $\tilde{A}_N(f)$ be the power spectra of the spacing and alignment noise and $\tilde{S}_N$ and $\tilde{A}_N$ be the total spacing and alignment noise powers at the output of a chain of N repeaters with an independent white timing noise of power N_0 added at the input of each repeater. From (27) and (67d),

$$\tilde{S}_N(f) = 4 \sin^2 \pi f \, \tilde{T}_N(f). \tag{116}$$

Substituting (105) and (93),

$$\tilde{S}_N(f) = N_0 \left\{ 4 \sinh^2 \frac{\pi}{2\tilde{Q}} [1 - | \tilde{H}(f) |^{2N}] \right\}. \tag{117}$$

From (112),

$$\tilde{S}_N = N_0 \left(\frac{\pi}{\tilde{Q}} \right)^2 \left(1 - \frac{1}{2\tilde{Q}} \sqrt{\frac{\pi}{N-1}} \right); \qquad \tilde{Q} \gg 1, \qquad N \gg 1. \tag{118}$$

The second term of (118) is negligible.

From (28),

$$\tilde{A}_N(f) = | \tilde{H}(f) - 1 |^2 [\tilde{T}_{N-1}(f) + N_0]. \tag{119}$$

From (91) and (93),

$$\operatorname{Re} \tilde{H}(f) = \frac{1 + e^{-\pi/\tilde{Q}}}{2} | \tilde{H}(f) |^2 + \frac{1 - e^{-\pi/\tilde{Q}}}{2}, \tag{120}$$

so that

$$| \tilde{H}(f) - 1 |^2 = e^{-\pi/\tilde{Q}}[1 - | \tilde{H}(f) |^2]. \tag{121}$$

Substituting (121) and (105) into (119),

$$\tilde{A}_N(f) = N_0 e^{-\pi/\tilde{Q}}[1 - | \tilde{H}(f) |^{2N}], \tag{122}$$

and from (112)

$$\tilde{A}_N = N_0 e^{-\pi/\tilde{Q}}\left[1 - \frac{1}{2\tilde{Q}} \sqrt{\frac{\pi}{N - 1}} \right]; \qquad \tilde{Q} \gg 1, \qquad N \gg 1. \tag{123}$$

The second term of (123) is again negligible.

Figs. 9(b) and 9(c) show $\tilde{S}_N(f)$ and $\tilde{A}_N(f)$. They have the same general shape, but the alignment noise has considerably greater magnitude. For high $\tilde{Q}$ they have an essentially white spectrum, except for very low frequencies; their spectra and total power change very little along the repeater chain. The main contribution to the spacing and alignment deviations comes from the high-frequency components of the input timing deviation, so only the input noise at the Nth repeater gives a significant contribution to these quantities.

The small angle restrictions of (21) were assumed to hold throughout the analysis. Equation (118) shows that if N_0 is small enough so that the first equation of (21) is satisfied for a single repeater, then the second equation of (21) will remain satisfied for the entire chain of repeaters.

4.4 Discussion

Each repeater acts as a discrete low-pass filter to the timing deviation, transmitting the low frequencies with little loss and attenuating the high-frequency components. The resulting timing noise at the output of a repeater chain, with noise added at every repeater, contains primarily low frequencies; the total timing noise power grows as the square root of the number of repeaters. The timing deviations of consecutive pulses are strongly correlated. Although the total timing deviation may become large, it changes so slowly that the spacing and alignment deviations remain small. The spacing and alignment noise have an almost white spectrum and remain almost constant along the repeater chain; successive spacing and alignment deviations are almost uncorrelated.

The results of this section are exact for locked-oscillator timing circuits and for tuned circuits with periodic pulse patterns containing every Mth pulse; for tuned circuits with random or general periodic pulse

patterns they are only approximate. The exact analysis for a repeater chain with a general periodic pulse pattern is discussed briefly in the following section; the approximate results of the present section are in good agreement with the exact results for this case. The present results also agree with those given by W. R. Bennett for the total output timing noise of a single repeater with a random pulse pattern.[7]

In order to apply these results to specific cases, a detailed analysis of the particular type of repeater is necessary to determine the power N_0 of the equivalent timing noise at the input of each repeater as a function of the real input noise, the pulse pattern, and other parameters of the system. Of course, N_0 will be proportional to the input noise power; its general dependence on the pulse pattern can be described in a qualitative way without a detailed analysis.[2]

Consider first tuned-circuit timing filters. From (115) and (90), the total output timing noise in a single repeater with an input white timing noise of power N_0 is approximately

$$\bar{\tau}_1 = N_0 \frac{\pi}{2pQ}; \qquad pQ \gg 1, \tag{124}$$

where Q is the tuned-circuit "Q" and p is the average number of time slots containing pulses. We consider four cases, discussed in the first section; in all of them from (50) the power of the timing wave is proportional to p^2.

1. Baseband repeater, signal pulses drive tuned circuit directly. The real output noise power remains constant for different pulse patterns; since $\bar{\tau}_1$ is proportional to the output noise to signal ratio, $\bar{\tau}_1 \propto 1/p^2$, and, from (124), $N_0 \propto 1/p$.

2. Baseband repeater, signal pulses passed through square-law or similar nonlinear device before driving tuned circuit. The real noise in the vacant time slots is now substantially suppressed; the real noise power at the output might be expected to vary approximately as p. Consequently $\bar{\tau}_1 \propto 1/p$, and N_0 is approximately independent of p.

3. Carrier frequency repeater, linear envelope detector. This case is somewhat different from case 1 above. During a time slot containing a pulse, only the in-phase component of the real input noise contributes to the random component of the envelope of the input wave. In the absence of a signal pulse, both the in-phase and quadrature components of the real input noise contribute equally to the output random phase modulation of the tuned circuit. Consequently, as p decreases, $\bar{\tau}_1$ will increase somewhat faster than $1/p^2$, and N_0 will increase somewhat faster than $1/p$.

4. Carrier frequency repeater, square-law or similar nonlinear detector. This type of repeater behaves essentially the same as a baseband repeater with a square-law device, case 2 above; N_0 is approximately independent of p. This has been verified in detail for the special case in which the timing circuit is an ideal flat filter, rather than a tuned circuit.

For locked oscillator timing circuits, (115) gives

$$\bar{\tau}_1 = N_0 \frac{\pi}{2\tilde{Q}}; \qquad \tilde{Q} \gg 1, \tag{125}$$

with $\tilde{Q}$, given by (87) or (88), now independent of the pulse pattern; the timing-wave amplitude is also constant. A similar discussion to the one above shows that N_0 for a locked oscillator has approximately the same functional dependence on p as it would for a tuned circuit for the corresponding type of repeater (i.e., baseband or carrier, linear or square-law detector). Since $\tilde{Q}$ is now independent of p, the variation of the output timing noise power $\bar{\tau}_1$ for a locked oscillator is p times that for a tuned circuit, with the same type of repeater.

Improved performance is thus attained by using a square-law or similar nonlinear detector in both baseband and carrier-frequency repeaters, using either tuned circuit or locked oscillator timing circuits, as pointed out by De Lange.[2] In this case the equivalent input timing noise power N_0 will be approximately independent of the pulse pattern, and may be treated as a constant in the analysis. Qualitatively, the principal effect of the input noise is simply to add a random displacement to the input signal pulses, as assumed in the analysis.

The above calculation of the alignment deviation must be modified for linear baseband repeaters (without a square-law or other nonlinear element) and for carrier frequency repeaters using a linear envelope detector. Here for $p < 1$ the equivalent position modulation required in the analysis will exceed the actual effective position modulation of the input signal pulses by the real input noise. Instead of (122) and (123) we would have approximately for the linear baseband repeater:

$$\tilde{A}_N(f) = N_0 \{ p e^{-\pi/\tilde{Q}} [1 - |\tilde{H}(f)|^{2N}] + (1 - p)|\tilde{H}(f)|^2 \}, \tag{126}$$

$$\tilde{A}_N = N_0 \left\{ p e^{-\pi/\tilde{Q}} \left(1 - \frac{1}{2\tilde{Q}} \sqrt{\frac{\pi}{N-1}} \right) + (1 - p)\frac{\pi}{2\tilde{Q}} \right\}; \tag{127}$$

$$\tilde{Q} \gg 1, \qquad N \gg 1,$$

with a similar result for the carrier repeater with a linear envelope detector. Since in this case $N_0 \propto 1/p$, the net effect as p decreases is an in-

crease in the low frequency portion of the alignment noise power spectrum, with only a small increase in the total alignment noise power. The over-all behavior of the alignment noise is thus very similar both for linear repeaters and for repeaters using a square-law or similar nonlinear element, for which (122) and (123) apply with N_0 a constant independent of p. The above results for timing and spacing noise, of course, remain valid for all cases.

V. EXACT ANALYSIS FOR A CHAIN OF REPEATERS USING TUNED-CIRCUIT TIMING FILTERS, FOR PERIODIC PULSE PATTERNS — ZERO TUNING ERROR, COMPLETE RETIMING

The results of Section IV are exact only in those cases where the timing wave amplitude $\tilde{A}_{n'}$ is strictly constant, i.e., for tuned circuits with a periodic pulse pattern with every Mth pulse present (including all pulses present), and for locked oscillators with any arbitrary pulse pattern. For tuned circuits with random or general periodic pulse patterns these results are only approximate. In this section the exact solution for a chain of repeaters using tuned circuit timing filters is discussed for a simple periodic pulse pattern; in this case the analysis of Section IV provides a very satisfactory approximation. As above, the repeaters are assumed to have zero tuning error and complete retiming, and the driving pulses for the tuned circuits are assumed to be short enough so that no additional timing noise results from the finite pulse width (Section 2.5). The response of a chain of repeaters to a single frequency input $e^{j2\pi f n'}$ first is determined and then used to determine the noise response of the system. Since the pulse pattern is the same, the variation of $\tilde{A}_{n'}$ is identical at every repeater; the repeaters will consequently be identical time-varying transducers to the timing deviation.

Setting $\delta f = 0$ in (35), which gives the exact response of a single repeater with a tuned-circuit timing filter, yields

$$\bar{\epsilon}^{\circ}(n') - \bar{\epsilon}^{\circ}(n' - 1) = \frac{1}{\tilde{A}_{n'}} [\bar{\epsilon}^{i}(n') - \bar{\epsilon}^{\circ}(n' - 1)]. \qquad (128)$$

The timing wave amplitude $\tilde{A}_{n'}$ is given by (24). For a general periodic pulse pattern $\tilde{A}_{n'}$ will be a periodic function of n'. The simplest case of interest is a periodic pulse pattern of period M containing only two pulses, located in arbitrary time slots; $\tilde{A}_{n'}$ then takes on only two values and may be written as

$$\tilde{A}_{n'} = A(1 + \delta e^{j\pi n'}), \qquad (129)$$

where A is the average timing wave amplitude and δ is the normalized deviation of $\tilde{A}_{n'}$ from its average value.

Equation (129) is the Fourier series for $\tilde{A}_{n'}$; A and δ are determined by (24). From (50),

$$A = \frac{2Q}{\pi M}; \qquad Q \gg \frac{\pi M}{2}; \tag{130}$$

δ will depend on the particular pulse pattern. The pattern giving the greatest variation in timing wave amplitude, and hence the largest δ, will consist of two adjacent pulses followed by $M - 2$ vacant time slots, as shown in Fig. 10. In this case,

$$\delta = -\frac{\pi}{4Q}(M - 2); \qquad Q \gg \frac{\pi M}{2}. \tag{131}$$

For other patterns containing two pulses, δ will vary between zero and the value given in (131). For periodic pulse patterns containing more pulses, the Fourier series for $\tilde{A}_{n'}$ corresponding to (129) will contain more terms.

We now seek the response of a repeater chain to the input $e^{j2\pi f n'}$. From (128) and (129) the input and output are related by a modulation process involving the term

$$e^{j\pi n'} = e^{j2\pi(1/2)n'}, \tag{132}$$

of frequency $\frac{1}{2}$. Since $e^{j2\pi n'} = 1$,

$$e^{j\pi n'} e^{j2\pi f n'} = e^{j2\pi(f-1/2)n'},$$
$$e^{j\pi n'} e^{j2\pi(f-1/2)n'} = e^{j2\pi f n'}. \tag{133}$$

Therefore, for $f > 0$, if the input timing error to a repeater contains a sinusoidal component at the frequency f the output will contain com-
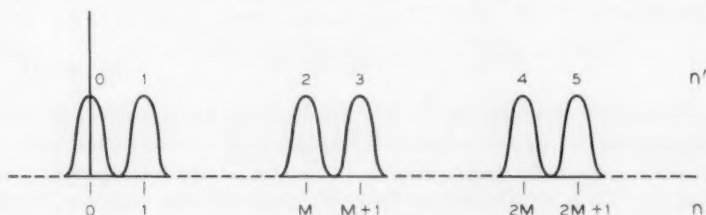
Fig. 10 — Periodic pulse pattern containing two pulses giving maximum variation in timing wave amplitude.

ponents at frequencies f and $f - \frac{1}{2}$; an input at the frequency $f - \frac{1}{2}$ will give rise to output components at frequencies $f - \frac{1}{2}$ and f. (If $f < 0$, an input at f yields outputs at f and $f + \frac{1}{2}$; an input at $f + \frac{1}{2}$ yields outputs at $f + \frac{1}{2}$ and f.) Assuming that $f > 0$, the timing error at the output of the $(K - 1)$th and the Kth repeaters may be written as follows:

$$\bar{\epsilon}_{K-1}{}^{o}(n') = A_{K-1}(f)e^{j2\pi f n'} + B_{K-1}(f)e^{j2\pi(f-1/2)n'}, \tag{134}$$

$$\bar{\epsilon}_{K}{}^{o}(n') = A_{K}(f)e^{j2\pi f n'} + B_{K}(f)e^{j2\pi(f-1/2)n'}. \tag{135}$$

Substituting (134) and (135) for $\bar{\epsilon}^{i}(n')$ and $\bar{\epsilon}^{o}(n')$ respectively in (128), and substituting (129) for $\bar{A}_{n'}$, equating the coefficients of each of the two frequencies on each side of the resulting equation, we obtain

$$aA_{K}(f) + bB_{K}(f) = A_{K-1}(f), \tag{136}$$

$$cA_{K}(f) + dB_{K}(f) = B_{K-1}(f), \tag{137}$$

where the parameters a, b, c, and d are constants which depend on the input frequency f in a simple way. Since the input to the repeater chain is $e^{j2\pi f n'}$, the initial conditions for the linear difference equations (136) and (137) are

$$A_{0} = 1, \qquad B_{0} = 0. \tag{138}$$

These difference equations are easily solved by the usual methods, yielding the response of the system to a single input frequency; the results may then be used to determine the timing, spacing and alignment noise along the repeater chain with an independent white timing noise added at the input of every repeater. Although the solution is straightforward the analysis is lengthy, and the problem does not seem to be of sufficiently general interest to warrant the inclusion of the detailed results. Examination of these results shows that for quite moderate values of Q the variation in the timing wave amplitude may be safely neglected in considering the timing behavior of a chain of repeaters. The approximate analysis of Section IV, which neglects the additional modulation products caused by the variation of the timing wave amplitude, is reliable for a long chain of repeaters with tuned-circuit timing filters for this particular pulse pattern.

This analysis may be generalized to more complicated periodic pulse patterns. However, as the number of pulses in one period increases, so does the number of simultaneous difference equations corresponding to (136) and (137), increasing the complexity of the analysis.

VI. THE EFFECT OF TUNING ERROR ON THE TIMING OF A CHAIN OF RE-PEATERS USING LOCKED-OSCILLATOR TIMING CIRCUITS WITH COMPLETE RETIMING

We next consider the timing deviations introduced by random tuning errors of the different repeaters, in a chain of repeaters using locked-oscillator timing circuits with complete retiming. From (37) the timing response of such a repeater is given by

$$\bar{\epsilon}^o(n') - \bar{\epsilon}^o(n'-1) = \frac{1}{A}\left[\bar{\epsilon}^i(n') - \bar{\epsilon}^o(n'-1)\right]$$

$$- \frac{\delta f}{F}\left(1 - \frac{1}{A}\right)b_{n'}, \tag{139}$$

where A is the ratio of the amplitude of the oscillator wave to the amplitude of the transient started by a single signal pulse. As pointed out in Section 2.3, this relation does *not* approximate the behavior of a tuned circuit with tuning error. Since no equivalent analysis has been devised for tuned circuits, the present treatment of the effects of tuning error on a repeater chain is confined exclusively to repeaters employing locked oscillators.

The output timing deviation of a repeater using a locked-oscillator timing circuit is linearly related to the input timing deviation, but has an additional equivalent input timing deviation related to the pulse pattern through the $b_{n'}$ and proportional to the tuning error. In the present section we consider only the timing noise caused by the tuning errors of the different repeaters, since the effects of input noise for this case have been discussed in Section IV.

We will consider only random pulse patterns with a probability p that any time slot contains a pulse, independently for each time slot. For such a pulse pattern the quantities $b_{n'}$ are independent random variables and consequently have a white spectrum, in addition to a dc component. The quantity $(b_{n'} - 1)$ has a geometric distribution,[14] so that the mean and variance of $b_{n'}$ are

$$\langle b_{n'} \rangle = \frac{1}{p},$$

$$\sigma_b^2 = \langle b_{n'}^2 \rangle - \langle b_{n'} \rangle^2 = \frac{1-p}{p^2}. \tag{140}$$

Separating $b_{n'}$ into dc and ac components,

$$b_{n'} = B + B_{n'}, \qquad B = \langle b_{n'} \rangle = \frac{1}{p}, \tag{141}$$

$$\langle B_{n'} \rangle = 0, \qquad \langle B_{n'}^2 \rangle = \frac{1 - p}{p^2}. \tag{142}$$

It is now convenient to separate (139) into dc and ac components, as in (40) to (45). Writing

$$\tilde{\epsilon}(n') = \tilde{\epsilon}_{dc} + \tilde{\epsilon}_{ac}(n'), \tag{143}$$

we have, from (139):

$$\tilde{\epsilon}_{dc}{}^i - \tilde{\epsilon}_{dc}{}^o = \frac{\delta f}{F}(A - 1)B, \tag{144}$$

$$\tilde{\epsilon}_{ac}{}^o(n') - \tilde{\epsilon}_{ac}{}^o(n' - 1) = \frac{1}{A}\left[\tilde{\epsilon}_{ac}{}^i(n') - \tilde{\epsilon}_{ac}{}^o(n' - 1)\right]$$

$$- \frac{\delta f}{F}\left(1 - \frac{1}{A}\right)B_{n'}. \tag{145}$$

Equation (144) represents a constant delay through the repeater due to tuning error. There will be a corresponding dc alignment deviation between the input signal and the retiming pulses, but otherwise the dc component is of no further interest in the analysis of the chain of repeaters. The ac timing deviation is governed by (145); for simplicity, the subscripts $_{ac}$ will be dropped in the remainder of this section.

We now let $\tilde{C}^i(f)$, $\tilde{C}^o(f)$ and $\tilde{B}(f)$ be the transforms of $\tilde{\epsilon}_{ac}{}^i(n')$, $\tilde{\epsilon}_{ac}{}^o(n')$ and $B_{n'}$ respectively. Taking the transform of (145),

$$\tilde{C}^o(f) = \tilde{H}(f)\left[\tilde{C}^i(f) - \frac{\delta f}{F}\frac{1}{e^{+\pi/\tilde{Q}} - 1}\tilde{B}(f)\right], \tag{146}$$

where

$$\tilde{H}(f) = \frac{1 - e^{-\pi/\tilde{Q}}}{1 - e^{-\pi/\tilde{Q}}e^{-j2\pi f}}, \tag{147}$$

$$A = \frac{1}{1 - e^{-\pi/\tilde{Q}}}; \qquad \tilde{Q} = \pi A, \qquad \tilde{Q} \gg 1. \tag{148}$$

$\tilde{H}(f)$ and $\tilde{Q}$ are the same transfer function and effective "Q" previously defined for the repeater in (91) and (87) and (88), and $B_{n'}$ is a white noise of power $(1 - p)/p^2$.

In Section IV the effects of independent white timing noise added at the input of every repeater were determined. The present analysis is somewhat different in that the added timing noises at the different repeater inputs are no longer independent, since the pulse pattern and hence $B_{n'}$ and $B(f)$ are identical at each repeater. In general, we no longer

may simply add the output power spectra caused by the various noise inputs, but must keep track of the phase angles of the different components. Thus the timing noise power spectrum $\tilde{T}_N(f)$ at the output of the Nth repeater, in terms of the tuning errors $(\delta f)_K$ of the various repeaters, becomes

$$\tilde{T}_N(f) =$$
$$\frac{1}{F^2}\left[\frac{1}{e^{+\pi/\tilde{Q}}-1}\right]^2\frac{1-p}{p^2}\sum_{K=1}^{N}\cdot\sum_{L=1}^{N}(\delta f)_{N-K+1}(\delta f)_{N-L+1}\tilde{H}^K(f)\tilde{H}^{*L}(f). \tag{149}$$

Next assume that the tuning errors $(\delta f)_K$ themselves are independent random variables with zero mean. Thus,

$$\langle(\delta f)_K(\delta f)_L\rangle = 0; \qquad K \neq L,$$
$$\langle(\delta f)_K^2\rangle = \langle(\delta f)^2\rangle; \qquad K = L. \tag{150}$$

The tuning errors at the different repeaters are uncorrelated. The equivalent input timing noises at the different repeaters are thus also uncorrelated, although not independent; in this special case, the output power spectra of the different components may be added directly. Substituting (150) into (149),

$$\tilde{T}_N(f) = \frac{\langle(\delta f)^2\rangle}{F^2}\left[\frac{1}{e^{+\pi/\tilde{Q}}-1}\right]^2\frac{1-p}{p^2}\frac{1-\mid\tilde{H}(f)\mid^{2N}}{\mid\tilde{H}(f)\mid^{-2}-1},$$
$$\langle(\delta f)_K(\delta f)_L\rangle = \begin{cases}0 & ; \quad K \neq L \\ \langle(\delta f)^2\rangle; & K = L\end{cases}. \tag{151}$$

Comparing with (105), the timing noise power spectrum caused by uncorrelated tuning errors has the same shape as that caused by independent white timing noise at every repeater input. Thus, the total output timing noise power in the present case is given by (111) or by (113) to (115), where N_0 must be replaced by

$$\frac{\langle(\delta f)^2\rangle}{F^2}\left[\frac{1}{e^{+\pi/\tilde{Q}}-1}\right]^2\frac{1-p}{p^2}.$$

For the high-$\tilde{Q}$ case:

$$\tilde{T}_1 = \frac{\langle(\delta f)^2\rangle}{F^2}\frac{\tilde{Q}}{\pi}\frac{1-p}{2p^2}; \qquad \tilde{Q} \gg 1,$$
$$\tilde{T}_N = \frac{\langle(\delta f)^2\rangle}{F^2}\frac{\tilde{Q}}{\pi}\frac{1-p}{p^2}\sqrt{\frac{N}{\pi}}; \qquad \tilde{Q} \gg 1, \quad N \gg 1. \tag{152}$$

The spacing and alignment noise power spectra and total powers are found in the same way from the results of Section 4.3.

The results of (151) and (152) are no longer valid if the tuning errors at the different repeaters are correlated. For example, if all tuning errors are identical, making use of (121), (149) becomes

$$\tilde{T}_N(f) = \left(\frac{\delta f}{F}\right)^2 \frac{e^{+\pi/\tilde{Q}}}{(e^{+\pi/\tilde{Q}} - 1)^2} \frac{1 - p}{p^2} \frac{|1 - \tilde{H}^N(f)|^2}{|\tilde{H}(f)|^{-2} - 1}, \tag{153}$$

$$(\delta f)_K = \delta f.$$

In this case, all repeaters are tuned correctly but the pulse repetition frequency is incorrect. The last factor of (153) approaches N^2 as f approaches zero [the corresponding factor of (151) or (105) approaches N, as illustrated in Fig. 9(a)]. The low-frequency noise is much greater in this case, since an identical timing noise has been added at every repeater and the corresponding noise amplitudes, rather than powers, add at the output.

For a single repeater using a locked oscillator, the output timing noise power given in (152) is just twice as large as the corresponding quantity computed by W. R. Bennett for a tuned circuit.[7] This is in accord with the discussion of the two cases given in (38) to (45) and illustrated in Fig. 6 for p close to 1. A similar analysis for the timing noise produced by tuning error in a chain of repeaters using tuned circuit timing filters has not been found.

VII. TIMING IN A CHAIN OF REPEATERS WITH PARTIAL RETIMING

The preceding analysis has been confined to the case of complete retiming, in which the timing deviation of each output signal pulse is equal to the timing deviation of the corresponding timing pulse. All of these results are easily extended to the case of partial retiming, in which the timing deviation of each output pulse depends linearly on the timing deviations of both timing and input signal pulses.

As stated in Section I, in systems employing complete retiming the timing pulses may be derived only from the input signal pulses. If such a system with timing from the output signal pulses could somehow be started, its subsequent timing behavior would be completely independent of the input pulse pattern. Thus, the repeater would have no way of determining when the timing pulses were properly centered on the input signal pulses. However, in systems employing partial retiming, the timing pulses may be derived from either the input or the output signal pulses.

Let $\bar{\epsilon}^i(n')$ and $\bar{\epsilon}^o(n')$ be the normalized timing deviations of the n'th input and output signal pulses with transforms $\tilde{C}^i(f)$ and $\tilde{C}^o(f)$, as before, and let $\bar{\epsilon}^t(n')$ be the timing deviation of the n'th timing pulse with transform $\tilde{C}^t(f)$. Then, following Pierce,[6] we assume that

$$\bar{\epsilon}^o(n') = \alpha\bar{\epsilon}^i(n') + (1 - \alpha)\bar{\epsilon}^t(n'),$$
$$\tilde{C}^o(f) = \alpha\tilde{C}^i(f) + (1 - \alpha)\tilde{C}^t(f). \tag{154}$$

For complete retiming assumed up to now, $\alpha = 0$; for no retiming, $\alpha = 1$.

Consider first a repeater with the timing wave derived from the input signal pulses. All of the previous analysis for a single repeater (Section II, Section 4.1 and the first part of Section VI) remains valid if we replace $\bar{\epsilon}^o(n')$ and $\tilde{C}^o(f)$ by $\bar{\epsilon}^t(n')$ and $\tilde{C}^t(f)$ respectively. Thus, from (86) to (91), for a repeater with zero tuning error

$$\frac{\tilde{C}^t(f)}{\tilde{C}^i(f)} = \tilde{H}(f) = \frac{1 - e^{-\pi/\tilde{Q}}}{1 - e^{-\pi/\tilde{Q}}e^{-j2\pi f}} , \tag{155}$$

where $\tilde{H}(f)$ and $\tilde{Q}$ remain the same as before. Defining $\tilde{H}_{\alpha i}(f)$ as the transfer function of a single repeater with partial retiming with timing derived from the input, from (154) and (155)

$$\tilde{H}_{\alpha i}(f) = \alpha + (1 - \alpha)\tilde{H}(f). \tag{156}$$

For $\alpha = 0$, $\tilde{H}_{\alpha i}(f) = \tilde{H}(f)$. Making use of (120),

$$| \tilde{H}_{\alpha i}(f) |^2 = \gamma + (1 - \gamma) | \tilde{H}(f) |^2,$$
$$\gamma = \alpha[1 - e^{-\pi/\tilde{Q}} + \alpha e^{-\pi/\tilde{Q}}], \tag{157}$$

where $| \tilde{H}(f) |^2$ is given in (93). Substituting $| \tilde{H}_{\alpha i}(f) |^2$ for $| \tilde{H}(f) |^2$ in (104) and (105), the timing noise power spectrum at the output of a chain of repeaters with an independent white timing noise added at the input to each repeater is determined as in Section 4.3. For moderate values of α the timing noise spectrum will be slightly larger at high frequencies than for complete retiming, $\alpha = 0$. Thus, the total output timing noise power will be slightly larger than for complete retiming, given in (110) to (115). The spacing and alignment noise power spectra may also be easily determined.

Next consider a repeater with the timing wave derived from the output signal pulses. For zero tuning error we now have

$$\frac{\tilde{C}^t(f)}{\tilde{C}^o(f)} = \tilde{H}(f) = \frac{1 - e^{-\pi/\tilde{Q}}}{1 - e^{-\pi/\tilde{Q}}e^{-j2\pi f}}. \tag{158}$$

Substituting into (154), the transfer function $\tilde{H}_{ao}(f)$ of a single repeater with partial retiming with timing derived from the output is

$$\tilde{H}_{ao}(f) = \frac{\alpha}{1 - (1 - \alpha)\tilde{H}(f)}. \tag{159}$$

The 3-db bandwidth of a single repeater for small α and high $\tilde{Q}$ is approximately

$$f_{3db} = \frac{\alpha}{2\tilde{Q}}; \qquad \alpha \ll 1, \qquad \tilde{Q} \gg 1. \tag{160}$$

The bandwidth is α times the bandwidth of the corresponding repeater with the timing wave derived from the input [which for small α will be only slightly larger than the value for complete retiming, given in (94)]. The output timing deviations are thus much smaller with timing from the output than with timing from the input (for zero tuning error). However, in the limit as $\alpha \to 0$ the behavior of the repeater becomes independent of the input pulses, and the effects of any noise sources in the regenerator itself become correspondingly more important.

The analysis of Section VI for a locked oscillator with tuning error is readily extended in a similar manner to a repeater with partial retiming. With timing from the input there is again little change in behavior for small α. However, with timing from the output the dc alignment deviation $\tilde{\epsilon}_{dc}{}^{i} - \tilde{\epsilon}_{dc}{}^{t}$ is proportional to $1/\alpha$. The bandwidth of the repeater decreases as above, while the equivalent input timing noise due to tuning error increases.

Both tuning error and additional noise sources in the regenerator will determine whether for a given small value of α any advantage can be obtained by deriving the timing wave from the output rather than the input. For complete retiming, $\alpha = 0$, the timing wave must be derived from the input.

VIII. DISCUSSION

To illustrate the above results we consider a chain of repeaters using resonant-circuit timing filters with zero tuning error, complete retiming (the timing wave must of course be derived from the input), driving pulses short enough so that no additional timing deviations are introduced by the finite pulse width and a random pulse pattern. We further assume that the equivalent normalized timing deviations added to the

input pulses are distributed uniformly between -0.1 and $+0.1$; the equivalent input timing noise power is then

$$N_0 = 0.0033, \qquad \sqrt{N_0} = 0.0578.$$

Take

$N = 100$ repeaters,

$Q = 100$, tuned circuit "Q",

$p = \frac{1}{2}$, probability that a time slot contains a pulse.

From (90)

$$\tilde{Q} = 50.$$

The rms values of the timing, spacing, and alignment noise at the output of the repeater chain, from (113), (118) and (123), are:

$$\sqrt{\tilde{T}_N} = 0.0344,$$
$$\sqrt{\tilde{S}_N} = 0.00363,$$
$$\sqrt{\tilde{A}_N} = 0.056.$$

All of these normalized quantities are of course measured in units of a pulse period.

Next consider the effect of finite pulse width. An extreme case may give some indication of the importance of this effect. Consider the change in output timing deviation caused by changing the pulse pattern from all pulses present to every Mth pulse present, with raised cosine driving pulses as illustrated in Fig. 7. From (63) and (33), the change in output timing deviation will be $w_s(\pi/Q)(M - 1)$. For overlapping pulses [Fig. 7(b)] and for $Q = 100$ as in the above example, from (64) and the table following it, $w_s = 0.1073$, and the change in output timing deviation will be $0.00337 (M - 1)$. Thus, for $Q = 100$ and overlapping pulses as shown in Fig. 7(b), changing the pulse pattern from all pulses present to every 10th pulse present causes a phase shift of $10.9°$ in each repeater.

The analysis for the effect of finite pulse width (or amplitude-to-phase conversion in the limiter of the timing system) in a chain of repeaters for a random pulse pattern is complicated by the fact that the timing deviation added at each repeater is directly related to the timing wave amplitude. It is not obvious that the variations in timing wave amplitude can be neglected in this case, as was done in studying the effects of input noise (which will be independent of the pulse pattern).

Finally, consider a chain of repeaters using locked-oscillator timing circuits with random tuning errors, complete retiming (with the timing

wave derived from the input), and a random pulse pattern. Assume

$$\tilde{Q} = 100,$$
$$p = \tfrac{1}{2}.$$

Then, from (151), the equivalent rms input timing noise per repeater is

$$45 \, \frac{\sqrt{\langle(\delta f)^2\rangle}}{F} \, .$$

The effects of tuning error will be the same as the effects of an input timing noise uniformly distributed between -0.1 and $+0.1$ (corresponding to $\sqrt{N_0} = 0.0578$, as in the first example) for an rms fractional tuning error of

$$\frac{\sqrt{\langle(\delta f)^2\rangle}}{F} = 0.0013.$$

As discussed above, similar results are not available for the effects of tuning error in a chain of repeaters using tuned circuit timing filters.

IX. ACKNOWLEDGMENTS

APPENDIX

Approximate Solution for $\tilde{T}_N$

In this appendix we present the approximate solution obtained by S. O. Rice for $\tilde{T}_N$, the normalized timing noise power at the output of a chain of N repeaters with an independent white timing noise of power N_0 introduced at the input to each repeater. This solution is most useful when the timing circuits have a high $\tilde{Q}$. From (109), (105) and (93),

$$\tilde{T}_N(f) = N_0 \frac{1 - |\tilde{H}(f)|^{2N}}{|\tilde{H}(f)|^{-2} - 1},$$

$$|\tilde{H}(f)|^2 = \frac{1}{1 + \operatorname{csch}^2 \dfrac{\pi}{2\tilde{Q}} \sin^2 \pi f}, \tag{161}$$

$$\tilde{T}_N = \int_{-1/2}^{1/2} \tilde{T}_N(f) \, df.$$

Substituting

$$\theta = 2\pi f \tag{162}$$

into (161), we obtain

$$\tilde{T}_N = \frac{N_0}{\pi} \int_0^\pi \frac{1 - y^N(\theta)}{y^{-1}(\theta) - 1} \, d\theta,$$

$$y(\theta) = \frac{1}{1 + a \sin^2 \dfrac{\theta}{2}}, \tag{163}$$

where we have replaced $\operatorname{csch}^2 \pi/(2\tilde{Q})$ by a to simplify the analysis. Making the following change of variable,

$$x = \tan \frac{\theta}{2}, \qquad \sin^2 \frac{\theta}{2} = \frac{x^2}{1 + x^2}, \qquad d\theta = \frac{2 \, dx}{1 + x^2}, \tag{164}$$

Equation (163) becomes

$$\tilde{T}_N = \frac{N_0}{\pi} \frac{2}{a} \int_0^\infty \frac{dx}{x^2} \left\{ 1 - \left[\frac{1 + x^2}{1 + (a + 1)x^2} \right]^N \right\}. \tag{165}$$

Integrating (165) by parts, we find

$$\tilde{T}_N = N_0 \frac{4N}{\pi} \int_0^\infty \frac{dx}{(1 + x^2)^2} \left[\frac{1 + x^2}{1 + (a + 1)x^2} \right]^{N+1}. \tag{166}$$

Making another change of variable

$$u = \frac{1 + (a + 1)x^2}{1 + x^2}, \qquad du = \frac{2ax \, dx}{(1 + x^2)^2}, \qquad x^2 = \frac{u - 1}{a + 1 - u}, \tag{167}$$

Equation (166) becomes

$$\tilde{T}_N = N_0 \frac{2N}{\pi a} \int_1^{a+1} \frac{du}{u^{N+1}} \left[\frac{(a + 1) - u}{u - 1} \right]^{1/2}. \tag{168}$$

Making one last change of variable,

$$v = u - 1, \tag{169}$$

Equation (168) becomes

$$\tilde{T}_N = N_0 \frac{2N}{\pi a} \int_0^a \frac{dv}{(1 + v)^{N+1}} \left(\frac{a - v}{v} \right)^{1/2}. \tag{170}$$

The principal contribution to the integral of (170) comes from the

singularity at $v = 0$. We split off this portion of the integral and use the remaining terms to get bounds on the error. Thus,

$$
\begin{aligned}
\tilde{T}_N &= N_0 \frac{2N}{\pi a} \left[\int_0^a \frac{dv}{v^{1/2}(1 + v)^{N+1}} a^{1/2} \right. \\
&\qquad \left. - \int_0^a \frac{dv}{v^{1/2}(1 + v)^{N+1}} [a^{1/2} - (a - v)^{1/2}] \right] \\
&= N_0 \frac{2N}{\pi a^{1/2}} \left\{ \int_0^\infty \frac{dv}{v^{1/2}(1 + v)^{N+1}} - \int_a^\infty \frac{dv}{v^{1/2}(1 + v)^{N+1}} \right. \\
&\qquad \left. - \int_0^a \frac{dv}{v^{1/2}(1 + v)^{N+1}} \left[\frac{a^{1/2} - (a - v)^{1/2}}{a^{1/2}} \right] \right\} \\
&= N_0 \frac{2N}{\pi a^{1/2}} \left[\frac{\Gamma(\tfrac{1}{2})\Gamma(N + \tfrac{1}{2})}{\Gamma(N + 1)} - R_1 - R_2 \right],
\end{aligned}
\tag{171}
$$

$$
R_1 = \int_a^\infty \frac{dv}{v^{1/2}(1 + v)^{N+1}},
$$

$$
R_2 = \int_0^a \frac{dv}{v^{1/2}(1 + v)^{N+1}} \left[\frac{a^{1/2} - (a - v)^{1/2}}{a^{1/2}} \right].
\tag{172}
$$

We next find bounds on R_1 and R_2.

For R_1, we have

$$
0 < R_1 = \int_a^\infty \frac{dv}{v^{1/2}(1 + v)^{N+1}} < \int_a^\infty \frac{dv}{v^{N+3/2}} = \frac{1}{a^{N+1/2}(N + \tfrac{1}{2})},
$$

$$
0 < R_1 < \frac{1}{a^{N+1/2}(N + \tfrac{1}{2})}.
\tag{173}
$$

For R_2, we first state the inequality

$$
a^{1/2} - (a - v)^{1/2} = \frac{v}{a^{1/2} + (a - v)^{1/2}} < \frac{v}{a^{1/2}}; \qquad 0 < v < a. \tag{174}
$$

Thus,

$$
\begin{aligned}
0 < R_2 &= \int_0^a \frac{dv}{v^{1/2}(1 + v)^{N+1}} \left[\frac{a^{1/2} - (a - v)^{1/2}}{a^{1/2}} \right] \\
&< \frac{1}{a} \int_0^a \frac{v^{1/2}\, dv}{(1 + v)^{N+1}} < \frac{1}{a} \int_0^\infty \frac{v^{1/2}\, dv}{(1 + v)^{N+1}} \\
&= \frac{1}{a} \frac{\Gamma(\tfrac{3}{2})\Gamma(N - \tfrac{1}{2})}{\Gamma(N + 1)},
\end{aligned}
\tag{175}
$$

$$
0 < R_2 < \frac{1}{a} \frac{\Gamma(\tfrac{3}{2})\Gamma(N - \tfrac{1}{2})}{\Gamma(N + 1)}.
$$

Replacing a by $\operatorname{csch}^2 \pi/2\tilde{Q}$ in (172), (173) and (175), we obtain as the final result

$$\tilde{T}_N = N_0 \frac{2N}{\pi} \sinh \frac{\pi}{2\tilde{Q}} \left[\frac{\Gamma(\frac{1}{2})\Gamma(N + \frac{1}{2})}{\Gamma(N + 1)} - R_1 - R_2 \right],$$

$$0 < R_1 < \sinh^{2N+1} \frac{\pi}{2\tilde{Q}} \left[\frac{1}{N + \frac{1}{2}} \right], \tag{176}$$

$$0 < R_2 < \sinh^2 \frac{\pi}{2\tilde{Q}} \left[\frac{\Gamma(\frac{3}{2})\Gamma(N - \frac{1}{2})}{\Gamma(N + 1)} \right].$$

The bounds are quite close for a moderately high $\tilde{Q}$, and improve with increasing $\tilde{Q}$ or N.

If $\tilde{Q} \gg 1$, $N \gg 1$, the sinh may be replaced by its argument and the gamma functions by Stirling's approximation, yielding the result given in (113):

$$\tilde{T}_N = N_0 \frac{\sqrt{\pi N}}{\tilde{Q}}; \qquad \tilde{Q} \gg 1, \qquad N \gg 1. \tag{177}$$

REFERENCES

1. DeLange, O. E., Experiments on the Regeneration of Binary Microwave Pulses, B.S.T.J., **35**, January 1956, pp. 67–91.
2. DeLange, O. E., this issue, pp. 1455–1486.
3. Sunde, E. D., Self-Timing Regenerative Repeaters, B.S.T.J., **36**, July 1957, pp. 891–937.
4. Wrathall, L. R., Transistorized Binary Pulse Regenerator, B.S.T.J. **35**, September 1956, pp. 1059–1084.
5. Bennett, W. R., Methods of Solving Noise Problems, Proc. I.R.E., **44**, May 1956, pp. 609–638.
6. Pierce, J. R., unpublished work.
7. Bennett, W. R., this issue, pp. 1501–1542.
8. Schelleng, J. C., unpublished work.
9. MacColl, L. A., *Fundamental Theory of Servomechanisms*, D. Van Nostrand Co., New York, 1945.
10. Tsien, H. S., *Engineering Cybernetics*, McGraw-Hill Book Co., New York, 1954.
11. James, H. M., Nichols, N. B. and Phillips, R. S., *Theory of Servomechanisms*, McGraw-Hill Book Co., New York, 1947.
12. Brown, W. M., Analysis of Discrete Linear Systems, J. Soc. Ind. & Appl. Math., **5**, December 1957, pp. 206–224.
13. Hildebrand, F. B., *Methods of Applied Mathematics*, Prentice-Hall, New York, 1952.
14. Feller, W., *Probability Theory and Its Applications*, John Wiley & Sons, New York, 1950.

Helix Waveguide Theory and Application

By HANS-GEORG UNGER

Generalized telegraphist's equations have been derived for curved helix waveguide and coefficients obtained for conversion from normal modes of the helix waveguide to normal modes of the metallic waveguide. A radial wave impedance at the helix interface is used to calculate the effect of composite jacket structures. Three different applications of the helix waveguide for circular electric wave transmission are discussed: As a mode filter, the helix waveguide should have a lossy jacket which causes a high transmission loss for all unwanted modes. For sharp intentional bends with tapered curvature, the helix waveguide should have a jacket of low-loss dielectric material surrounded by a highly conducting coaxial shield. For an all-helix waveguide, in order to reduce both mode conversion-reconversion effects at imperfections and loss in curvature the jacket should be medium lossy and also surrounded by a metallic shield. The distance between helix and shield should in all applications be about a quarter of the radial wavelength in the material. Measurements on unwanted mode transmission and TE_{01} curvature loss confirm the analysis.

I. INTRODUCTION

Helix waveguide consisting of closely wound insulated copper wire covered with a jacket of dielectric material and surrounded by a coaxial metallic shield has been shown to be useful as a communication medium.[1] The low-pitch helix gives to this structure the low-loss characteristics of solid copper-wall waveguide for circular-electric waves. All other modes of propagation, however, suffer a change of their field configuration and propagation constants, depending on the properties of the outside jacket and shield.[2]

There are two distinctive applications of the helix waveguide for circular electric wave transmission, requiring different but definite properties of jacket and shield. One is the use of helix waveguide in short sections inserted as mode filters in an otherwise conventional waveguide, to reduce the mode conversion-reconversion distortion of TE_{01} propaga-

tion. For this application the attenuation constant of all unwanted modes should be made as high as possible by choosing a material with a certain loss factor for the jacket. The other application is to transmit the circular-electric wave around sharp intentional bends.[3] In this case the jacket of the helix waveguide should be as low in loss as possible and the distance between helix and outer metallic shield should be chosen so as to minimize mode conversion and dissipation loss in bends.

Beyond this, a third and probably the most important application is the all-helix waveguide as transmission medium. In this case, high unwanted mode attenuation as well as low TE_{01} loss in bends is desired. A compromise between the helix waveguide for mode filters and the helix waveguide for sharp bends has to be met by choosing a medium lossy jacket and the proper distance between jacket and shield. This compromise will be made on the basis of tolerances. While the manufacturing tolerances call for a mode-filtering helix waveguide, the laying tolerances call for low TE_{01} loss in random curvature.

In order to design these different types of helix waveguide, we must analyze the propagation characteristics of both the circular electric mode and the unwanted modes in the general structure of helix surrounded by lossy jacket and metallic shield with both a straight and a curved axis. In the following two sections two different approaches to this problem will be presented. Both of them make use of Schelkunoff's generalized telegraphist's equations for waveguides.[4] The propagation in the curved helix waveguide is represented by a set of coupled modes of propagation. But the set of modes used in the first approach is different from the one used in the second approach. In Sections II and III the generalized telegraphist's equations are derived. In Section IV the problem of a composite jacket structure is reduced to a radial transmission line problem. The theory is applied to practical helix waveguide problems in Section V, and conclusions are actually given there. The reader interested only in design and application of helix waveguide is therefore referred immediately to Sections IV and V.

II. REPRESENTATION IN TERMS OF NORMAL MODES OF THE METALLIC WAVEGUIDE

To obtain generalized telegraphist's equations the field at any cross section of the curved helix waveguide is represented as a superposition of the fields of a certain set of normal modes. A current and a voltage amplitude are associated with each normal mode and the currents and voltages, after Maxwell's equations and the boundary conditions of the curved helix waveguide are applied, are found to satisfy an infinite set

of generalized telegraphist's equations. The coupling terms in these equations depend upon the curvature of the guide axis and the jacket structure. The choice of normal modes may be arbitrary, but, in this case, it seems most appropriate to choose the normal modes of a straight circular waveguide with perfectly conducting walls. It appears that, with this choice of modes, the generalized telegraphist's equations will also solve the problem of mode conversion at an abrupt or gradual transition from metallic waveguide to helix waveguide. Also, it will be possible to analyze mode conversion at all those imperfections of the helix waveguide, for which the mode conversion in a metallic waveguide is known.

The natural coordinate system (r, φ, z) for a curved circular waveguide is toroidal (Fig. 1), where z is the distance measured along the curved axis of the guide and r and φ are polar coordinates in a plane normal to the axis of the guide, with origin at the guide axis. The lines $\varphi = 0$ and $\varphi = \pi$ lie in the plane of the bend. The inner radius of the guide is denoted by a and the radius of the bend by R.

For the moment, (r, φ, z) are regarded as general orthogonal curvilinear coordinates (u, v, w) by letting

$$u = r \qquad v = \varphi \qquad w = z. \tag{1}$$

The element of length in this system is

$$ds^2 = e_1^2 du^2 + e_2^2 dv^2 + e_3^2 dw^2, \tag{2}$$

where

$$e_1 = 1; \qquad e_2 = r; \qquad e_3 = 1 + \xi \tag{3}$$

and

$$\xi = \frac{r}{R} \cos \varphi. \tag{4}$$

For a field with time dependence $e^{j\omega t}$, Maxwell's equations are:

$$\frac{1}{e_2 e_3} \left[\frac{\partial}{\partial v} (e_3 E_w) - \frac{\partial}{\partial w} (e_2 E_v) \right] = -j\omega\mu H_u, \tag{5}$$

$$\frac{1}{e_3 e_1} \left[\frac{\partial}{\partial w} (e_1 E_u) - \frac{\partial}{\partial u} (e_3 E_w) \right] = -j\omega\mu H_v, \tag{6}$$

$$\frac{1}{e_1 e_2} \left[\frac{\partial}{\partial u} (e_2 E_v) - \frac{\partial}{\partial v} (e_1 E_u) \right] = -j\omega\mu H_w, \tag{7}$$

$$\frac{1}{e_2 e_3} \left[\frac{\partial}{\partial v} (e_3 H_w) - \frac{\partial}{\partial w} (e_2 H_v) \right] = j\omega\epsilon E_u , \tag{8}$$

$$\frac{1}{e_3 e_1} \left[\frac{\partial}{\partial w} (e_1 H_u) - \frac{\partial}{\partial u} (e_3 H_w) \right] = j\omega\epsilon E_v , \tag{9}$$

$$\frac{1}{e_1 e_2} \left[\frac{\partial}{\partial u} (e_2 H_v) - \frac{\partial}{\partial v} (e_1 H_u) \right] = j\omega\epsilon E_w . \tag{10}$$

The permeability μ and the permittivity ϵ are those of free space in the waveguide interior, but between helix and shield ϵ is determined by the choice of the jacket. If there is dissipation in the jacket it is complex.

To convert Maxwell's equations into generalized telegraphist's equations we introduce the field components of the normal modes of our choice. Each mode is described by a transverse field distribution pattern $T(u, v)$, which satisfies

$$\nabla^2 T = \frac{1}{e_1 e_2} \left[\frac{\partial}{\partial u} \left(\frac{e_2}{e_1} \frac{\partial T}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{e_1}{e_2} \frac{\partial T}{\partial v} \right) \right] = -\chi^2 T, \tag{11}$$

where χ is a separation constant which has discrete values for the various TE and TM modes. Brackets will be used for TE modes and parentheses for TM modes. Thus, the function corresponding to the nth TE mode is denoted by $T_{[n]}(u, v)$ and the separation constant by $\chi_{[n]}$ with the subscript in brackets. The function corresponding to the nth TM mode is denoted by $T_{(n)}(u, v)$ and the separation constant by $\chi_{(n)}$, with the subscript in parenthesis. The normal derivative of $T_{[n]}$ vanishes on the boundary $u = a$, while $T_{(n)}$ itself vanishes at $u = a$.

The T functions are assumed to be so normalized that

$$\int_S (\text{grad } T)(\text{grad } T) \, dS$$
$$= \int_S (\text{flux } T)(\text{flux } T) \, dS = \chi^2 \int_S T^2 \, dS = 1, \tag{12}$$

where S is the cross section of the guide inside the helix.

With the definition of the gradient and flux of T:

$$\text{grad}_u \, T = \frac{\partial T}{e_1 \partial u}, \qquad \text{grad}_v \, T = \frac{\partial T}{e_2 \partial v},$$

$$\text{flux}_u \, T = \frac{\partial T}{e_2 \partial v}, \qquad \text{flux}_v \, T = -\frac{\partial T}{e_1 \partial u}, \tag{13}$$

various orthogonality relations exist among the T functions:

$$\int_s T_{(n)} T_{(m)} \, dS = \int_s T_{[n]} T_{[m]} \, dS = 0,$$

$$\int_s (\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS = \int_s (\text{flux } T_{(n)})(\text{flux } T_{(m)}) \, dS = 0, \quad (14)$$

$$\int_s (\text{grad } T_{[n]})(\text{grad } T_{[m]}) \, dS = \int_s (\text{flux } T_{[n]})(\text{flux } T_{[m]}) \, dS = 0$$

if $m \neq n$, and

$$\int_s (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS = \int_s (\text{grad } T_{[n]})(\text{flux } T_{(m)}) \, dS$$

$$= \int_s (\text{grad } T_{(n)})(\text{flux } T_{(m)}) \, dS = 0 \tag{15}$$

for all m and n. We now assume series expansions of the field components in terms of the functions $T(u, v)$, with coefficients I and V depending on w:

$$E_u = \sum_n \left[V_{(n)} \frac{\partial T_{(n)}}{e_1 \partial u} + V_{[n]} \frac{\partial T_{[n]}}{e_2 \partial v} \right],$$

$$E_v = \sum_n \left[V_{(n)} \frac{\partial T_{(n)}}{e_2 \partial v} - V_{[n]} \frac{\partial T_{[n]}}{e_1 \partial u} \right],$$

$$H_u = \sum_n \left[-I_{(n)} \frac{\partial T_{(n)}}{e_2 \partial v} + I_{[n]} \frac{\partial T_{[n]}}{e_1 \partial u} \right], \tag{16}$$

$$H_v = \sum_n \left[I_{(n)} \frac{\partial T_{(n)}}{e_1 \partial u} + I_{[n]} \frac{\partial T_{[n]}}{e_2 \partial v} \right].$$

With the T functions, the field components in (16) are only defined inside of the helix for $u < a$. The effects of jacket and shield (at radius b) are taken into account by the boundary conditions:

$$E_v = 0, \tag{17}$$

$$E_w = -ZH_v \tag{18}$$

at the helix. In these conditions, the low-pitch helix has been replaced by an anisotropically conducting sheath, and jacket and shield by a wall impedance Z. Later on in this paper we shall discuss how to calculate this wall impedance and even how to take the effect of finite wire size of the helix into account by modifying Z.

The boundary condition (17) is satisfied by the individual terms of the series for E_v.

To transform Maxwell's equations into generalized telegraphist's equations the series expansions (16) are substituted for the field components in (5) through (10). Then certain combinations of these equations are integrated over the cross section and advantage is taken of the orthogonality relation (14) and (15). For example, substituting from (16) into (5) and (6), adding

$$-\frac{e_3}{e_2}\frac{\partial T_{(m)}}{\partial v} \text{ times (5)} \qquad \text{and} \qquad \frac{e_3}{e_1}\frac{\partial T_{(m)}}{\partial u} \text{ times (6)}$$

and integrating over the cross section, we obtain:

$$\frac{dV_{(m)}}{dw} + j\omega\mu \sum_n \left[I_{(n)} \int_S e_3 (\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS \right.$$

$$\left. + I_{[n]} \int_S e_3 (\text{flux } T_{[n]})(\text{grad } T_{(m)}) \, dS \right] \qquad (19)$$

$$= \int_S (\text{grad } e_3 E_w)(\text{grad } T_{(m)}) \, dS.$$

The right-hand side can be integrated by parts according to Green's theorem:

$$a \int_0^{2\pi} e_3 E_w \frac{\partial T_{(m)}}{\partial u} \bigg|_a \, dv - \int_S e_3 E_w \text{ div grad } T_{(m)} \, dS.$$

In the first term of this expression, E_w is replaced by H_v through the boundary condition (18) and, for H_v in turn, the series expansion of (16) is substituted:

$$a \int_0^{2\pi} e_3 E_w \frac{\partial T_{(m)}}{\partial u} \bigg|_a \, dv = -aZ \sum_n \left[I_{(n)} \int_0^{2\pi} \frac{e_3}{e_1} \frac{\partial T_{(n)}}{\partial u} \frac{\partial T_{(m)}}{\partial u} \, dv \right.$$

$$\left. + I_{[n]} \int_0^{2\pi} \frac{e_3}{e_2} \frac{\partial T_{[n]}}{\partial v} \frac{\partial T_{(m)}}{\partial u} \, dv \right].$$

In the second term of the foregoing expression, E_w is replaced by H_v and H_u through (10) and, for H_v and H_u in turn, the series expansion of (16) is substituted:

$$j\omega\epsilon E_w = - \sum_n I_{(n)} \chi_{(n)}^2 T_{(n)}. \qquad (20)$$

Therefore:

$$-\int_S e_3 E_w \text{ div grad } T_{(m)} \, dS = \frac{j}{\omega\epsilon} \sum_n I_{(n)} \chi_{(n)}^2 \chi_{(m)}^2 \int_S e_3 T_{(n)} T_{(m)} \, dS.$$

Thus we get, instead of (19)

$$
\begin{aligned}
\frac{dV_{(m)}}{dw} = & -j\omega\mu \sum_n I_{(n)} \left[\int_S e_3(\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS \right. \\
& -\frac{\chi_{(n)}^2 \chi_{(m)}^2}{\omega^2 \mu \epsilon} \int_S e_3 T_{(n)} T_{(m)} \, dS \\
& \left. +\frac{aZ}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_1} \frac{\partial T_{(n)}}{\partial u} \frac{\partial T_{(m)}}{\partial u} \, dv \right] \\
& -j\omega\mu \sum_n I_{[n]} \left[\int_S e_3 \, (\text{flux } T_{[n]})(\text{grad } T_{(m)}) \, dS \right. \\
& \left. +\frac{aZ}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_2} \frac{\partial T_{[n]}}{\partial v} \frac{\partial T_{(m)}}{\partial u} \, dv \right].
\end{aligned}
\tag{21}
$$

By again substituting from (16) into (5) and (6), adding

$$
\frac{e_3}{e_1} \frac{\partial T_{[m]}}{\partial u} \text{ times (5)} \qquad \text{and} \qquad \frac{e_3}{e_2} \frac{\partial T_{[m]}}{\partial v} \text{ times (6)}
$$

and integrating over the cross section we obtain:

$$
\begin{aligned}
\frac{dV_{[m]}}{dw} + j\omega\mu \sum_n I_{[n]} \int_S e_3 \, (\text{grad } T_{[n]})(\text{grad } T_{[m]}) \, dS & \\
+ j\omega\mu \sum_n I_{(n)} \int_S e_3 \, (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS & \\
= -\int_S (\text{flux } e_3 E_w)(\text{grad } T_{[m]}) \, dS. &
\end{aligned}
\tag{22}
$$

Integrating the right-hand side of (22) by parts:

$$
\int_0^{2\pi} e_3 E_w \frac{\partial T_{[m]}}{\partial v} \bigg|_a \, dv,
$$

and replacing E_w by H_v through the boundary condition (18), (22) becomes:

$$
\begin{aligned}
\frac{dV_{[m]}}{dw} = & -j\omega\mu \sum_n I_{[n]} \left[\int_S e_3 \, (\text{grad } T_{[n]})(\text{grad } T_{[m]}) \, dS \right. \\
& \left. +\frac{Z}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_2} \frac{\partial T_{[n]}}{\partial v} \frac{\partial T_{[m]}}{\partial v} \, dv \right] \\
& -j\omega\mu \sum_n I_{(n)} \left[\int_S e_3 \, (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS \right. \\
& \left. +\frac{Z}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_1} \frac{\partial T_{(n)}}{\partial u} \frac{\partial T_{[m]}}{\partial v} \, dv \right].
\end{aligned}
\tag{23}
$$

Similarly, by adding

$$-\frac{e_3}{e_1}\frac{\partial T_{(m)}}{\partial u} \text{ times (8)} \qquad \text{and} \qquad -\frac{e_3}{e_2}\frac{\partial T_{(m)}}{\partial v} \text{ times (9),}$$

substituting from (7) for H_w and integrating over the cross section we get:

$$\frac{dI_{(m)}}{dw} = -j\omega\epsilon \sum_n \left[V_{(n)} \int_S e_3 (\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS \right.$$
$$\left. + V_{[n]} \int_S e_3 (\text{grad } T_{(m)})(\text{flux } T_{[n]}) \, dS \right]. \tag{24}$$

Finally, by adding

$$-\frac{e_3}{e_2}\frac{\partial T_{[m]}}{\partial v} \text{ times (8)} \qquad \text{and} \qquad \frac{e_3}{e_1}\frac{\partial T_{[m]}}{\partial u} \text{ times (9),}$$

again substituting from (7) for H_w and integrating over the cross section we get:

$$\frac{dI_{[m]}}{dw} = -j\omega\epsilon \sum_n V_{[n]} \left[\int_S e_3 (\text{grad } T_{[n]})(\text{grad } T_{[m]}) dS \right.$$
$$\left. - \frac{\chi_{[n]}^2 \chi_{[m]}^2}{\omega^2 \mu\epsilon} \int_S e_3 T_{[n]} T_{[m]} \, dS \right] \tag{25}$$
$$- j\omega\epsilon \sum_n V_{(n)} \int_S e_3 (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS.$$

Equations (21), (23), (24) and (25) are generalized telegraphist's equations for the curved helix waveguide. They can be written in the following form:

$$\frac{dV_{(m)}}{dw} = -\sum_n [Z_{(m)(n)}I_{(n)} + Z_{(m)[n]}I_{[n]}],$$

$$\frac{dV_{[m]}}{dw} = -\sum_n [Z_{[m](n)}I_{(n)} + Z_{[m][n]}I_{[n]}],$$

$$\tag{26}$$

$$\frac{dI_{(m)}}{dw} = -\sum_n [Y_{(m)(n)}V_{(n)} + Y_{(m)[n]}V_{[n]}],$$

$$\frac{dI_{[m]}}{dw} = -\sum_n [Y_{[m](n)}V_{(n)} + Y_{[m][n]}V_{[n]}].$$

The impedance and admittance coefficients are defined by

$$Z_{(m)(n)} = j\omega\mu \left[\int_S e_3 \, (\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS \right.$$
$$\left. - \frac{\chi_{(n)}^2 \chi_{(m)}^2}{\omega^2 \mu \epsilon} \int_S e_3 T_{(n)} T_{(m)} \, dS + \frac{aZ}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_1} \frac{\partial T_{(n)}}{\partial u} \cdot \frac{\partial T_{(m)}}{\partial u} \, dv \right],$$

$$Z_{(m)[n]} = j\omega\mu \left[\int_S e_3 \, (\text{grad } T_{(m)})(\text{flux } T_{[n]}) \, dS \right.$$
$$\left. + \frac{aZ}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_2} \frac{\partial T_{[n]}}{\partial v} \frac{\partial T_{(m)}}{\partial u} \, dv \right],$$

$$Z_{[m](n)} = j\omega\mu \left[\int_S e_3 \, (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS \right.$$
$$\left. + \frac{Z}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_1} \frac{\partial T_{(n)}}{\partial u} \frac{\partial T_{[m]}}{\partial v} \, dv \right],$$

$$Z_{[m][n]} = j\omega\mu \left[\int_S e_3 \, (\text{grad } T_{[n]})(\text{grad } T_{[m]}) \, dS \right. \tag{27}$$
$$\left. + \frac{Z}{j\omega\mu} \int_0^{2\pi} \frac{e_3}{e_2} \frac{\partial T_{[n]}}{\partial v} \frac{\partial T_{[m]}}{\partial v} \, dv \right],$$

$$Y_{(m)(n)} = j\omega\epsilon \int_S e_3 \, (\text{grad } T_{(n)})(\text{grad } T_{(m)}) \, dS,$$

$$Y_{(m)[n]} = j\omega\epsilon \int_S e_3 \, (\text{grad } T_{(m)})(\text{flux } T_{[n]}) \, dS,$$

$$Y_{[m](n)} = j\omega\epsilon \int_S e_3 \, (\text{grad } T_{(n)})(\text{flux } T_{[m]}) \, dS,$$

$$Y_{[m][n]} = j\omega\epsilon \left[\int_S e_3 \, (\text{grad } T_{[n]})(\text{grad } T_{[m]}) \, dS \right.$$
$$\left. - \frac{\chi_{[n]}^2 \chi_{[m]}^2}{\omega^2 \mu \epsilon} \int_S e_3 T_{[n]} T_{[m]} \, dS \right].$$

We note the following symmetry properties of these coefficients:

$$Z_{(m)(n)} = Z_{(n)(m)} , \qquad Y_{(m)(n)} = Y_{(n)(m)} ;$$
$$Z_{[m][n]} = Z_{[n][m]} , \qquad Y_{[m][n]} = Y_{[n][m]} ; \tag{28}$$
$$Z_{[m](n)} = Z_{(n)[m]} , \qquad Y_{[m](n)} = Y_{(n)[m]} .$$

The generalized telegraphist's equations represent an infinite set of

mutually coupled transmission lines. For our purposes, it is more convenient to write the transmission-line equations not in terms of currents and voltages but in terms of the amplitudes of forward and backward travelling waves. Thus, let a and b be the amplitudes of the forward and backward waves of a typical mode at a certain cross section. The mode current and voltages are related to the wave amplitudes a and b by

$$V = \sqrt{K}\,(a + b),$$
$$I = \frac{1}{\sqrt{K}}\,(a - b), \tag{29}$$

where K is the wave impedance

$$K_{(n)} = \frac{\beta_{(n)}}{\omega\epsilon}, \qquad K_{[n]} = \frac{\omega\mu}{\beta_{[n]}},$$
$$\beta_{(n)} = (\beta^2 - \chi_{(n)}{}^2)^{1/2}, \qquad \beta_{[n]} = (\beta^2 - \chi_{[n]}{}^2)^{1/2}, \qquad \beta^2 = \omega^2\mu\epsilon.$$

The currents and voltages in the generalized telegraphist's equations (26) are represented in terms of the travelling-wave amplitudes. The following equations for coupled travelling waves are obtained after some obvious additions and subtractions:

$$\frac{da_{(m)}}{dz} = -\sum_n [\kappa_{(m)(n)}{}^+ a_{(n)} + \kappa_{(m)(n)}{}^- b_{(n)} + \kappa_{(m)[n]}{}^+ a_{[n]} + \kappa_{(m)[n]}{}^- b_{[n]}],$$

$$\frac{db_{(m)}}{dz} = +\sum_n [\kappa_{(m)(n)}{}^- a_{(n)} + \kappa_{(m)(n)}{}^+ b_{(n)} + \kappa_{(m)[n]}{}^- a_{[n]} + \kappa_{(m)[n]}{}^+ b_{[n]}],$$

$$\frac{da_{[m]}}{dz} = -\sum_n [\kappa_{[m](n)}{}^+ a_{(n)} + \kappa_{[m](n)}{}^- b_{(n)} + \kappa_{[m][n]}{}^+ a_{[n]} + \kappa_{[m][n]}{}^- b_{[n]}], \tag{30}$$

$$\frac{db_{[m]}}{dz} = +\sum_n [\kappa_{[m](n)}{}^- a_{(n)} + \kappa_{[m](n)}{}^+ b_{(n)} + \kappa_{[m][n]}{}^- a_{[n]} + \kappa_{[m][n]}{}^+ b_{[n]}].$$

The κ's are coupling coefficients defined by

$$\kappa_{(m)(n)}{}^{\pm} = \frac{1}{2}\left[\sqrt{K_{(m)}K_{(n)}}\; Y_{(m)(n)} \pm \frac{1}{\sqrt{K_{(m)}K_{(n)}}}\; Z_{(m)(n)}\right],$$

$$\kappa_{(m)[n]}{}^{\pm} = \frac{1}{2}\left[\sqrt{K_{(m)}K_{[n]}}\; Y_{(m)[n]} \pm \frac{1}{\sqrt{K_{(m)}K_{[n]}}}\; Z_{(m)[n]}\right],$$

$$\kappa_{[m](n)}{}^{\pm} = \frac{1}{2}\left[\sqrt{K_{[m]}K_{(n)}}\; Y_{[m](n)} \pm \frac{1}{\sqrt{K_{[m]}K_{(n)}}}\; Z_{[m](n)}\right], \tag{31}$$

$$\kappa_{[m][n]}{}^{\pm} = \frac{1}{2}\left[\sqrt{K_{[m]}K_{[n]}}\; Y_{[m][n]} \pm \frac{1}{\sqrt{K_{[m]}K_{[n]}}}\; Z_{[m][n]}\right].$$

In these definitions the plus signs are taken together, as are the minus signs.

To examine the coupling coefficients, we introduce field functions of the normal modes of the circular guide with perfectly conducting walls. We shall use the customary double-subscript notation, but shall continue to denote TM waves with parentheses and TE waves with brackets:

$$
\begin{aligned}
T_{(nm)} &= \sqrt{\frac{\epsilon_n}{\pi}} \frac{J_n(\chi_{(nm)}r)\sin n\varphi}{k_{(nm)}J_{n-1}(k_{nm})}, \\
T_{[nm]} &= \sqrt{\frac{\epsilon_n}{\pi}} \frac{J_n(\chi_{[nm]}r)\cos n\varphi}{(k_{[nm]}{}^2 - n^2)^{1/2}J_n(k_{nm})},
\end{aligned}
\tag{32}
$$

where

$$
\begin{aligned}
k_{(nm)} &= \chi_{(nm)}a, & J_n(k_{(nm)}) &= 0; \\
k_{[nm]} &= \chi_{[nm]}a, & J_n'(k_{[nm]}) &= 0
\end{aligned}
\tag{33}
$$

and

$$
\begin{aligned}
\epsilon_n &= 1, & n &= 0, \\
\epsilon_n &= 2, & n &\neq 0.
\end{aligned}
\tag{34}
$$

Introducing (32) into (27) and these in turn into (31), the coupling coefficients are calculated.

There are two different causes of coupling between the metallic guide normal modes. One is the finite wall impedance Z of the helix waveguide; we shall continue to denote the corresponding coupling coefficient with κ. The other is the curvature of the guide axis; we shall denote the corresponding coupling coefficient with jc. The coefficients of curvature coupling jc turn out to be purely imaginary.

The κ's in (31) which have equal subscripts may be regarded as propagation constants of typical TM or TE modes which have been modified by the finite wall impedance Z. It turns out that these modified propagation constants do not depend on the curvature:

$$
\begin{aligned}
\kappa_{(nm)(nm)} &= \gamma_{(nm)} = j\beta_{(nm)} + \frac{\epsilon_n}{2a}\frac{Z}{K_{(nm)}}, \\
\kappa_{[nm][nm]} &= \gamma_{[nm]} = j\beta_{[nm]} + \frac{\epsilon_n}{2a}\frac{n^2}{k_{[nm]}{}^2 - n^2}\frac{Z}{K_{[nm]}}.
\end{aligned}
\tag{35}
$$

For all other κ's we find that the wall impedance causes coupling only

between modes of equal subscript n in (32). The corresponding coupling coefficients are:

$$K_{(nm)(np)}{}^+ = \frac{\epsilon_n}{2a} \frac{Z}{\sqrt{K_{(nm)}K_{(np)}}},$$

$$K_{(nm)[np]}{}^+ = K_{[np](nm)}{}^+ = -\frac{\epsilon_n}{2a} \frac{n}{\sqrt{k_{[np]}{}^2 - n^2}} \frac{Z}{\sqrt{K_{(nm)}K_{[np]}}}, \qquad (36)$$

$$K_{[nm][np]}{}^+ = \frac{\epsilon_n}{2a} \frac{n^2}{\sqrt{k_{[nm]}{}^2 - n^2}\sqrt{k_{[np]}{}^2 - n^2}} \frac{Z}{\sqrt{K_{[nm]}K_{[np]}}}.$$

The curvature of the guide axis causes coupling only between modes which have a subscript n in (32) differing by one. The corresponding coupling coefficients are the same as in the curved circular guide with perfectly conducting walls.[5] We obtain, for example,

$$c_{[01](11)}{}^+ = \frac{\beta}{\sqrt{2}\,k_{[01]}} \frac{a}{R},$$

$$c_{[01](1m)}{}^+ = 0 \qquad \text{for } m \neq 1,$$

$$c_{[01][1m]}{}^+ = \frac{\sqrt{2}\,k_{[01]}k_{[1m]}{}^2}{\sqrt{k^2_{[1m]} - 1}(k^2_{[01]} - k^2_{[1m]})^2} \left(\frac{\sqrt{2}\,\beta^2 a^2 - k^2_{[01]} - k^2_{[1m]}}{a^2\,\sqrt{2\beta_{[01]}\beta_{[1m]}}} \right. \qquad (37)$$

$$\left. + \sqrt{\beta_{[01]}\beta_{[1m]}} \right) \frac{a}{R}.$$

The curvature coupling between TE_{01} and all other modes does not depend on the finite wall impedance of the helix waveguide.

There is some combined curvature and wall-impedance coupling between certain higher order modes. However, for TE_{01} transmission in the helix waveguide, this is of no interest.

To apply the preceding analysis to practical problems we need consider only the forward waves, since the relative power coupled from the forward waves into the backward waves is quite small. After the b's and κ^-'s have been omitted, (30) can conveniently be written in matrix form:

$$A' = -MA. \qquad (38)$$

If we designate the amplitude of the TE_{01} by a_0 and let the remaining a's represent the amplitudes of the modes which are coupled to TE_{01} in

the curved helix waveguide in order of their cut-off frequencies, then, in a somewhat simplified notation, the matrices in (38) are defined by:

$$A' = \frac{dA}{dz}, \qquad A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \qquad M = \begin{bmatrix} j\beta_0 & jc_{01} & jc_{02} & \cdots \\ jc_{01} & \kappa_{11} & \kappa_{12} & \cdots \\ jc_{02} & \kappa_{12} & \kappa_{22} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}. \quad (39)$$

If both the c's and off-diagonal κ's were small, the solution of (38) could be written down as a perturbation of the propagation in the straight metallic waveguide. This is usually not the case. In gentle bends such as we are interested in here, however, the curvature coupling is orders of magnitude smaller than the wall impedance coupling between the modes. Therefore, it is convenient to split M into $M_h + M_c$, where M_h is the straight helix waveguide matrix and M_c takes the curvature effects into account:

$$M_h = \begin{bmatrix} j\beta_0 & 0 & 0 & \cdots \\ 0 & \kappa_{11} & \kappa_{12} & \cdots \\ 0 & \kappa_{12} & \kappa_{22} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}, \qquad M_c = \begin{bmatrix} 0 & jc_{01} & jc_{02} & \cdots \\ jc_{01} & 0 & 0 & \cdots \\ jc_{02} & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}. \quad (40)$$

The effect of curvature can be calculated by a perturbation method, but

$$A' = M_h A \qquad (41)$$

will have to be solved more rigorously. To do this, we diagonalize M_h by the transformation

$$A = LW, \qquad (42)$$

where L is the modal matrix[6] of M_h. Since M_h is symmetrical, we have $L_t = L^{-1}$ and the transformation of M_h to the diagonal form of its latent roots[6] reads:

$$L_t M_h L = \Gamma.$$

In

$$\Gamma = \begin{bmatrix} \gamma_0 & 0 & 0 & \cdots \\ 0 & \gamma_1 & 0 & \cdots \\ 0 & 0 & \gamma_2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix} \tag{43}$$

the γ's are solutions of the characteristic equation

$$\begin{vmatrix} j\beta_0 - \gamma & 0 & 0 & \cdots \\ 0 & \kappa_{11} - \gamma & \kappa_{12} & \cdots \\ 0 & \kappa_{12} & \kappa_{22} - \gamma & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{vmatrix} = 0. \tag{44}$$

A typical element of the modal matrix L of M_h is defined by:

$$l_{ks} = \frac{\overset{s}{q}_{ik}}{\sqrt{\left(\overset{s}{q}_{i1}\right)^2 + \left(\overset{s}{q}_{i2}\right)^2 + \cdots \left(\overset{s}{q}_{in}\right)^2}}, \tag{45}$$

where $\overset{s}{q}_{ik}$ is the cofactor of an element of the determinant in (44) for the root $\gamma = \gamma_s$. The index i is arbitrary as long as it is different from zero but it must, of course, be the same for the determination of all the l's for a given s.

Physically, (42) means transformation to the normal modes of the straight helix waveguide, and the γ's are the propagation constants of the normal modes in the straight helix waveguide. The elements of L in (42) describe the conversion of normal modes of the straight helix waveguide into normal modes of the metallic waveguide. The TE_{01} amplitude and propagation constants are, of course, not changed by this transformation:

$$w_0 = a_0 , \qquad \gamma_0 = j\beta_0 .$$

Introducing the normal modes w into (38), we get for the propagation in the curved helix waveguide

$$W' = -(\Gamma + C)W, \tag{46}$$

where $C = L_t M_c L$ represents the effect of curvature in the helix wave-guide. The elements of C are the coefficients of coupling between the normal modes of the straight helix waveguide in curved sections, and C is a symmetrical matrix with nonvanishing elements only in the first row and first column. A typical element is given by

$$c_s = j(c_{01} l_{1s} + c_{02} l_{2s} + c_{03} l_{32} + \cdots). \tag{47}$$

Since Γ in (46) is diagonal, and since the elements of C are small enough to justify a perturbation calculation, the problem of the curved helix waveguide is formally solved.

The perturbation calculation for the curved guide assumes, as usual, that the coupling from TE_{01} to all other modes w is small. Then it is sufficient to consider only coupling between TE_{01} and one of the other modes at a time and to add up all these coupling effects. The TE_{01} propagation in a helix waveguide, even of nonuniform curvature, may be calculated this way.[3]

For a helix waveguide of uniform curvature our matrix notation may be used. As before with the straight helix waveguide, we now transform to the normal modes of the curved helix waveguide. The propagation constants of these curved-guide normal modes are the solution of the characteristic equation:

$$\begin{vmatrix} \gamma - \gamma_0 & c_1 & c_2 & \cdots \\ c_1 & \gamma - \gamma_1 & 0 & \cdots \\ c_2 & 0 & \gamma - \gamma_2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{vmatrix} = 0. \tag{48}$$

Approximate solutions of this equation for

$$\left| \frac{c_s}{\gamma_s - \gamma_s} \right| \ll 1$$

are perturbations of the propagation constants γ_s of normal modes of the straight helix waveguide. In particular, we get for the perturbed propagation constant of the TE_{01} mode:

$$\gamma = \gamma_0 + \sum_s \frac{c_s^2}{\gamma_0 - \gamma_s}. \tag{49}$$

The modal matrix of $(\Gamma + C)$ transforms from the normal modes of the curved guide to those of the straight guide. Its elements give the

mode conversion at the junction from curved to straight guide. A typical element of this modal matrix is given by an expression like (45) or to a sufficient approximation:

$$l_{0s} = \frac{c_s}{\gamma_0 - \gamma_s}. \tag{50}$$

The propagation constants γ_s in (49) and (50) are those of the normal modes in the helix waveguide. To obtain them, the characteristic equation (44), a polynomial of infinite degree in γ, has to be solved.

For a low wall-impedance jacket, approximate solutions of (44) may be used which, like (49), are perturbations of the metallic waveguide propagation constants. In general, (44) may be reduced to a polynomial of finite degree by omitting modes of higher order which have little effect on the lower order modes. Even so, we are usually left with the formidable problem of solving a polynomial of high degree.

To avoid the problem associated with the transformation from metallic guide normal modes to helix guide normal modes we must start with the helix waveguide normal modes and write the generalized telegraphist's equations in terms of normal modes of the helix waveguide.

III. REPRESENTATION IN TERMS OF NORMAL MODES OF THE HELIX WAVE-GUIDE

Normal modes of the straight helix waveguide have been analyzed elsewhere.[2] To adapt this analysis to our representation we shall repeat and generalize the boundary value problem here. The waveguide structure we consider is shown in Fig. 1. The anisotropic conducting sheath,
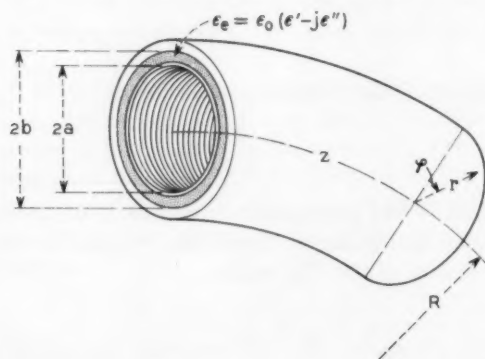
Fig. 1 — Coordinates used in bend in shielded helix waveguide.

which represents the closely wound helix, is surrounded by a dielectric jacket and a coaxial metallic shield.

The electromagnetic field in the helix waveguide can be derived from two sets of scalar functions T_n and T_n' given by

$$T_n = N_n J_p(\chi_n r) \sin p\varphi$$
$$T_n' = N_n J_p(\chi_n r) \cos p\varphi \qquad \text{for } 0 < r < a, \qquad (51a)$$

and

$$T_n = N_n \frac{\chi_n^2}{\chi_n^{e^2}} J_p(k_n) \frac{H_p^{(2)}(\chi_n^e r) - c H_p^{(1)}(\chi_n^e r)}{H_p^{(2)}(k_n^e) - c H_p^{(1)}(k_n^e)} \sin p\varphi$$

$$\text{for } a < r < b. \quad (51b)$$

$$T_n' = N_n \frac{\chi_n^2}{\chi_n^{e^2}} J_p(k_n) \frac{H_p^{(2)}(\chi_n^e r) - c' H_p^{(1)}(\chi_n^e r)}{H_p^{(2)}(k_n^e) - c' H_p^{(1)}(k_n^e)} \cos p\varphi$$

Being solutions of the wave equation, the T functions satisfy (11). The field components are written in terms of these field functions, using the coordinates (u, v, w) of (1):

$$E_u = \sum_n V_n \left[\frac{\partial T_n}{e_1 \partial u} + d_n \frac{\partial T_n'}{e_2 \partial v} \right],$$

$$E_v = \sum_n V_n \left[\frac{\partial T_n}{e_2 \partial v} - d_n \frac{\partial T_n'}{e_1 \partial u} \right],$$

$$H_u = \sum_n -I_n \left[\frac{\partial T_n}{e_2 \partial v} - d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_1 \partial u} \right] \frac{\epsilon}{\epsilon_0}, \qquad (52)$$

$$H_v = \sum_n I_n \left[\frac{\partial T_n}{e_1 \partial u} + d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_2 \partial v} \right] \frac{\epsilon}{\epsilon_0}.$$

Substituting from (52) into (7) and (10) and taking advantage of (11), we get for the longitudinal field components:

$$H_w = j\omega\epsilon \sum_n V_n d_n \frac{\chi_n^2}{k^2} T_n',$$

$$E_w = j\omega\mu \sum_n I_n \frac{\epsilon}{\epsilon_0} \frac{\chi_n^2}{k^2} T_n, \qquad (53)$$

where ϵ and $k = \omega\sqrt{\mu\epsilon}$ are permittivity and intrinsic propagation constant of the medium in a particular cross-sectional part of the guide. They have constant but different values for the different cross-sectional parts of the guide, and ϵ_0 is the permittivity of the empty helix interior.

The quantities d_n, c and c' and the separation constants χ_n and $\chi_n{}^e$ are chosen so that the boundary conditions of the shielded helix waveguide

$$
\begin{aligned}
E_v &= 0 & &\text{at helix and shield,} \\
E_w &= 0 & &\text{at shield,} \\
E_{w\text{interior}} &= E_{w\text{exterior}} & &\text{at helix,} \\
H_{v\text{interior}} &= H_{v\text{exterior}} & &\text{at helix}
\end{aligned}
\tag{54}
$$

are satisfied by the individual terms of (52). Only then do the individual terms of (52) represent normal modes of the shielded helix waveguide. From $E_v = 0$ at the helix:

$$
d_n = \frac{\dfrac{\partial T_n}{e_2 \partial v}}{\dfrac{\partial T_n{}'}{e_1 \partial u}}\Bigg|_{u=a} .
\tag{55}
$$

From $E_w = 0$ at the shield:

$$
c = \frac{H_p{}^{(2)}(\rho k_n{}^e)}{H_p{}^{(1)}(\rho k_n{}^e)} .
\tag{56}
$$

From $E_v = 0$ at the shield:

$$
c' = \frac{H_p{}^{(2)\prime}(\rho k_n{}^e)}{H_p{}^{(1)\prime}(\rho k_n{}^e)} ,
\tag{57}
$$

where $\rho = b/a$, $k_n{}^e = \chi_n{}^e a$, and $k_n = \chi_n a$. The prime at the Bessel functions denotes differentiation with respect to the argument. The condition of E_w being continuous across the helix boundary is satisfied by virtue of the formulation of the T functions in (51). The remaining condition of H_v being continuous across the boundary leads to the following (characteristic) equation of the shielded helix waveguide:

$$
\frac{J_p{}'(k_n)}{J_p(k_n)} - \frac{p^2}{k_n{}^2}\frac{h_n{}^2}{k^2}\frac{J_p(k_n)}{J_p{}'(k_n)} = \frac{\epsilon_e}{\epsilon_0}\frac{k_n}{k_n{}^e}\left[\frac{H_p{}^{(2)\prime}(k_n{}^e) - cH_p{}^{(1)\prime}(k_n{}^e)}{H_p{}^{(2)}(k_n{}^e) - cH_p{}^{(1)}(k_n{}^e)}\right.
$$
$$
\left. - \frac{p^2}{k_n{}^{e^2}}\frac{h_n{}^2}{k_e{}^2}\frac{H_p{}^{(2)}(k_n{}^e) - c'H_p{}^{(1)}(k_n{}^e)}{H_p{}^{(2)\prime}(k_n{}^e) - c'H_p{}^{(1)\prime}(k_n{}^e)}\right].
\tag{58}
$$

The characteristic equation, together with

$$
\begin{aligned}
k_n{}^2 &= (k^2 - h_n{}^2)a^2, \\
k_n{}^{e^2} &= (k_e{}^2 - h_n{}^2)a^2
\end{aligned}
\tag{59}
$$

determines the separation constants k_n and k_n^e. The transverse field components of any two different modes are orthogonal to each other in that

$$
\frac{1}{V_n I_m} \int_S (E_{tn} \times H_{tm}) \, dS
$$

$$
= \int_S \frac{\epsilon}{\epsilon_0} \left[\left(\frac{\partial T_n}{e_1 \partial u} + d_n \frac{\partial T_n'}{e_2 \partial v} \right) \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_2 \partial v} \right) \right.
$$
$$
\left. + \left(\frac{\partial T_n}{e_2 \partial v} - d_n \frac{\partial T_n'}{e_1 \partial u} \right) \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_1 \partial u} \right) \right] dS \tag{60}
$$

$$
= \delta_{nm} .
$$

The integration is to be extended over the entire cross section. The quantity δ_{nm} is the Kronecker delta. To satisfy (60) for $n = m$ requires the normalization N_n to have a certain value. Both (60) and N_n are calculated in the Appendix.

We have now determined all quantities in (51) and (52) except the current and voltage coefficients. To find relations for them we substitute (52) for the field components into Maxwell's equations and convert to generalized telegraphist's equations. We add

$$
-e_3 \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_1 \partial u} \right) \frac{\epsilon}{\epsilon_0} \text{ times (5)}
$$

and

$$
e_3 \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_2 \partial v} \right) \frac{\epsilon}{\epsilon_0} \text{ times (6)}
$$

and integrate over the cross section. Using (11) and (60) and the boundary conditions of (54), the result is:

$$
\frac{dV_m}{dw} + j \frac{h_m^2}{\omega \epsilon_0} I_m = -j\omega\mu \sum_n I_n \left\{ \int_S \xi \frac{\epsilon^2}{\epsilon_0^2} \left[\left(\frac{\partial T_n}{e_1 \partial u} \right. \right. \right.
$$
$$
+ d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_2 \partial v} \right) \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_2 \partial v} \right)
$$
$$
+ \left(\frac{\partial T_n}{e_2 \partial v} - d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_1 \partial u} \right) \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_1 \partial u} \right) \right] dS \tag{61}
$$
$$
- \int_S \xi \frac{\epsilon}{\epsilon_0} \frac{\chi_n^2 \chi_m^2}{k^2} T_n T_m \, dS \bigg\} .
$$

Similarly, we add

$$-e_3 \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{\partial T_m{}'}{e_2 \partial v} \right) \text{ times (8)}$$

and

$$-e_3 \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{\partial T_m{}'}{e_1 \partial u} \right) \text{ times (9)}$$

and integrate over the cross section:

$$
\begin{aligned}
\frac{dI_m}{dw} + j\omega\epsilon_0 V_m &= -j\omega\epsilon_0 \sum_n V_n \left\{ \int_s \xi \frac{\epsilon}{\epsilon_0} \left[\left(\frac{\partial T_n}{e_1 \partial u} \right. \right. \right. \\
&+ d_n \frac{\partial T_n{}'}{e_2 \partial v} \bigg) \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{\partial T_m{}'}{e_2 \partial v} \right) \\
&+ \left(\frac{\partial T_n}{e_2 \partial v} - d_n \frac{\partial T_n{}'}{e_1 \partial u} \right)\left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{\partial T_m{}'}{e_1 \partial u} \right) \bigg] dS \\
&- \int_s \xi \frac{\epsilon}{\epsilon_0} d_n d_m \frac{\chi_n^2 \chi_m^2}{k^2} T_n{}' T_m{}' \, dS \bigg\}.
\end{aligned}
\tag{62}
$$

Equations (61) and (62) are the generalized telegraphist's equations for the curved helix waveguide. The specification of the normalization factor in (60), which seemed rather arbitrary at that time, turns out in (61) and (62) to be the only right one. Only with this normalization are the mutual impedance and admittance coefficients symmetrical:

$$Y_{nm} \cong Y_{mn},$$

$$Z_{nm} = Z_{mn},$$

as we must expect from the symmetry properties of the curved helix waveguide structure.

Introducing travelling waves

$$V_m = \sqrt{K_m} \, (a_m + b_m),$$

$$I_m = \frac{1}{\sqrt{K_m}} \, (a_m - b_m)$$

$$\tag{63}$$

into (61) and (62) with $K_m = h_m/\omega\epsilon_0$, we get the more convenient form:

$$\frac{da_m}{dw} + jh_m a_m = j \sum_n (c_{mn}{}^+ a_n + c_{mn}{}^- b_n),$$

$$\frac{db_m}{dw} + jh_m b_m = -j \sum_n (c_{mn}{}^+ b_n + c_{mn}{}^- a_n).$$

$$\tag{64}$$

The coupling coefficients in (64) are:

$$c_{mn}^{\pm} =$$

$$\mp \frac{k^2}{2\sqrt{h_m h_n}} \int_S \xi \frac{\epsilon^2}{\epsilon_0^2} \left[\left(\frac{\partial T_n}{e_1 \partial u} + d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_2 \partial v} \right) \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_2 \partial v} \right) \right.$$

$$+ \left. \left(\frac{\partial T_n}{e_2 \partial v} - d_n \frac{h_n^2}{k^2} \frac{\partial T_n'}{e_1 \partial u} \right) \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{h_m^2}{k^2} \frac{\partial T_m'}{e_1 \partial u} \right) \right] dS$$

$$- \frac{1}{2} \sqrt{h_m h_n} \int_S \xi \frac{\epsilon}{\epsilon_0} \left[\left(\frac{\partial T_n}{e_1 \partial u} + d_n \frac{\partial T_n'}{e_2 \partial v} \right) \left(\frac{\partial T_m}{e_1 \partial u} + d_m \frac{\partial T_m'}{e_2 \partial v} \right) \right. \tag{65}$$

$$+ \left. \left(\frac{\partial T_n}{e_2 \partial v} - d_n \frac{\partial T_n'}{e_1 \partial u} \right) \left(\frac{\partial T_m}{e_2 \partial v} - d_m \frac{\partial T_m'}{e_1 \partial u} \right) \right] dS$$

$$\pm \frac{k^2}{2\sqrt{h_m h_n}} \int_S \xi \frac{\epsilon}{\epsilon_0} \frac{\chi_n^2 \chi_m^2}{k^2} T_n T_m \, dS$$

$$+ \frac{1}{2} \sqrt{h_m h_n} \int_S \xi \frac{\epsilon}{\epsilon_0} d_n d_m \frac{\chi_n^2 \chi_m^2}{k^2} T_n' T_m' \, dS.$$

We are interested here only in coupling between circular electric and other waves. Let the subscript m refer to a TE_{0m} wave. Then $J_1(k_m)$ is 0, and

$$T_m = 0, \qquad T_m' = N_m J_0(\chi_m r).$$

From (60) we get as normalization factor:

$$d_m N_m = \frac{1}{\sqrt{\pi}} \frac{k}{h_m} \frac{1}{k_m J_0(k_m)}.$$

Evaluation of (65) then yields:

$$c_{mn}^{\pm} = N_n \frac{\sqrt{\pi}}{2ka} \sqrt{\frac{h_n}{h_m}} \frac{k_m k_n^2}{k_m^2 - k_n^2} J_1(k_n) \left[1 \pm \frac{h_m}{h_n} \right.$$

$$+ \left. \left(\frac{h_m + h_n}{h_m - h_n} \right)^{\pm 1} \frac{J_1(k_n)}{k_n J_1'(k_n)} \right] \frac{1}{R}$$

$$= \frac{1}{\sqrt{2}} \sqrt{\frac{h_n}{h_m}} \frac{1}{ka} \frac{k_m k_n^2}{k_m^2 - k_n^2} \tag{66}$$

$$\cdot \frac{1 \pm \dfrac{h_m}{h_n} + \left(\dfrac{h_m + h_n}{h_m - h_n} \right)^{\pm 1} Y_n}{\left[\dfrac{h_n^2}{k^2} (k_n^2 - 1) Y_n^2 + \dfrac{1}{Y_n^2} + k_n^2 \left(1 - \dfrac{1}{k^2 a^2} \right) \right.} \cdot \frac{1}{R},$$

$$\left. + 2 \left(\dfrac{1}{Y_n} - Y_n \right) + \dfrac{\epsilon_e}{\epsilon_0} \dfrac{k_n^4}{k_n^{e4}} \left(2 k_n^e \cot \delta k_n^e + \dfrac{\delta k_n^{e3}}{\sin^2 \delta k_n^e} \right) \right]^{1/2}$$

where

$$Y_n = \frac{J_1(k_n)}{k_n J_1'(k_n)}$$

and

$$\delta = \rho - 1.$$

With (64) and (66), the circular electric wave propagation in the curved helix waveguide can be analyzed. Neglecting backward travelling waves in (64) and considering only coupling between a circular electric wave and one other mode at a time, the coupling coefficients $c_{mn}{}^+$ and propagation constants h_n can be used in formulae like (49) and (50) to compute circular electric wave loss and mode conversion in the curved helix waveguide.

A theory of the helix waveguide, however, is complete only when the transmission and conversion properties of waves at a transition from metallic guide to helix guide is known. When we represented helix waveguide propagation in terms of normal modes of the metallic guide, we obtained these transmission and conversion properties with the matrix L in (42). To obtain the corresponding matrix in the present formulation, we express a typical normal mode of the helix waveguide in terms of normal modes of the metallic guide. To do this exactly, an infinite set of linear equations in an infinite number of unknowns would have to be solved. However, as before we may safely neglect reflected waves and match only one pair of field components. The problem is then considerably simplified, especially when we choose the longitudinal field components to be matched.

The longitudinal field components of a typical forward mode in the helix waveguide are

$$E_{zn} = j\omega\mu \sqrt{\frac{\omega\epsilon}{h_n} \frac{\chi_n^2}{k^2}} \, T_n a_n ,$$

$$H_{zn} = j\omega\epsilon \sqrt{\frac{h_n}{\omega\epsilon} \frac{\chi_n^2}{k^2}} \, d_n T_n' a_n. \tag{67}$$

Longitudinal field components of forward modes in the metallic waveguide are

$$E_{z(n)} = j\omega\mu \sqrt{\frac{\omega\epsilon}{\beta_{(n)}} \frac{\chi_{(n)}^2}{k^2}} \, T_{(n)} a_{(n)} ,$$

$$H_{z[n]} = j\omega\epsilon \sqrt{\frac{\omega\mu}{\beta_{[n]}} \frac{\chi_{[n]}^2}{k^2}} \, T_{[n]} a_{[n]} . \tag{68}$$

Matching the field components:

$$E_{zn} = \sum_n E_{z(n)} , \qquad H_{zn} = \sum_n H_{z[n]}$$

we can, by means of a Fourier-Bessel expansion, express the wave amplitudes of the modes in the metallic waveguide in terms of the wave amplitudes of the modes in the helix waveguide:

$$a_{(m)} = l_{(m)n}a_n ,$$
$$a_{[m]} = l_{[m]n}a_n . \tag{69}$$

The coefficients $l_{(m)n}$ and $l_{[m]n}$ are the elements of a matrix L which corresponds, in our present formulation, to the matrix L of (42).

To improve the approximation, we actually did not match the longitudinal field components. Rather, we obtained two slightly different sets of l's by matching first the transverse components of the electric field and then the transverse components of the magnetic field. We expect that, if we would determine a set of l's from an exact solution of the problem, the values would lie somewhere between the results of these two calculations. Since the two approximate results are near together for modes sufficiently far from cutoff, their geometric mean should furnish a still better approximation to the exact l. Typical coefficients l are thus given by:

$$l_{(m)n} = \chi_n^{\ 2} \int_S T_n T_{(m)} \, dS,$$
$$l_{[m]n} = \frac{\beta_{[m]}}{k} \chi_n^{\ 2} d_n \int_S T_n' T_{[m]} \, dS. \tag{70}$$

The integrals in (70) are extended only over the part of the cross section which is inside the helix.

Calculation of the reciprocal coefficients proved the law of reciprocity $l_{n(m)} = l_{(m)n}$ and $l_{n[m]} = l_{[m]n}$ to be satisfied when the normalization factor N_n is chosen according to (60).

Evaluation of the integrals in (70) yields:

$$l_{(m)n} = \sqrt{\epsilon_p \pi} \, N_n \frac{k_n^{\ 2} J_p(k_n)}{k_n^{\ 2} - k_{(pm)}^{\ 2}},$$
$$l_{[m]n} = \frac{\sqrt{\epsilon_p \pi} \, p N_n}{\sqrt{k_{[pm]}^{\ 2} - p^2}} \frac{\beta_{[pm]}}{k} \frac{k_n^{\ 2} J_p(k_n)}{k_n^{\ 2} - k_{[pm]}^{\ 2}}, \tag{71}$$

where the expression for N_n is given in (91) of the Appendix. There is only conversion between modes of equal azimuthal order p.

IV. MODIFIED JACKET STRUCTURES

The helix waveguide which has been analyzed in the preceding sections has a wall structure of very special and highly idealized form. Practical jackets may intentionally or unintentionally be quite different from this mathematical model. We shall discuss several modifications of the ideal shielded helix waveguide which have proven to be important in a practical helix waveguide.[7] They are either changes of the ideal structure to simplify the manufacturing process or improve the performance, or they are deviations from the ideal structure, which are unavoidable in a real structure. It turns out that, for most of these modifications, we need not work out a new theory (as in the preceding sections) but can include their effects in the foregoing formulae. A quantity which is very useful for these discussions is the wall impedance of the helix waveguide:

4.1 *The Wall Impedance of the Helix Waveguide*

In all cases of practical interest the jacket permittivity is high enough so that

$$| k_n^e | \gg | (4p^2 - 1)/8 |$$

is well satisfied and the Hankel functions may be replaced by their asymptotic expressions. The characteristic equation (58) for the normal modes of the helix waveguide then reduces to:

$$\frac{1}{k_n} \frac{J_p{}'(k_n)}{J_p(k_n)} - \frac{p^2 h_n{}^2}{k_n{}^3 k^2} \frac{J_p(k_n)}{J_p{}'(k_n)} = -\frac{\epsilon_e}{\epsilon_0} \frac{1}{k_n^e} \cot \delta k_n^e. \tag{72}$$

The right-hand side of (72) can, to the same approximation, be expressed by the impedance

$$Z = -\frac{E_{wn}}{H_{vn}} = j \frac{\chi_n^e}{\omega \epsilon_e} \tan \delta k_n^e, \tag{73}$$

which the wall presents to a typical mode of the helix waveguide:

$$\frac{1}{k_n} \frac{J_p{}'(k_n)}{J_p(k_n)} - \frac{p^2 h_n{}^2}{k_n{}^3 k^2} \frac{J_p(k_n)}{J_p{}'(k_n)} = \frac{-j}{\omega \epsilon_0 a Z}. \tag{74}$$

Thus, the separation constant k_n and with it the helix waveguide mode is determined by the wall impedance alone.

It is, on the other hand, quite easy to determine the wall impedance of a composite jacket structure to the same approximation. The physical meaning of the approximations which have been made to derive (73) is $\delta \ll 1$. The thickness of the dielectric jacket or the penetration of waves

into it is small enough so that the coaxial structure can be considered plane in cartesian coordinates. Furthermore, the dependence on the azimuthal coordinate φ or v respectively is of so low an order compared to the r or u and z or w dependence that it may be neglected. The wall impedance we then calculate and quite generally use in the characteristic equation (74) is that of a symmetrical TM wave at a plane interface. For these waves the wall impedance of (73) is an exact representation of the boundary condition.

4.2 Unshielded Helix Waveguide

When the lossy jacket surrounding the helix is either so thick or has so high a loss factor that the electromagnetic field has died away before it reaches the metallic shield, we may remove it or substitute a different supporting structure without affecting the electrical performance of the helix waveguide. Some formulae in Section III may be simplified in this case, by letting b, the radius of the shield, go to infinity. For example, the characteristic equation (58) for the normal modes of the helix waveguide reduces to:

$$\frac{J_p{'}(k_n)}{J_p(k_n)} - \frac{p^2 h_n^{\,2}}{k_n^{\,2} k^2} \frac{J_p(k_n)}{J_p{'}(k_n)} = \frac{\epsilon_e}{\epsilon_0} \frac{k_n}{k_n^{\,e}} \left[\frac{H_p^{(2)'}(k_n^{\,e})}{H_p^{(2)}(k_n^{\,e})} - \frac{p^2}{k_n^{\,e2}} \frac{h_n^{\,2}}{k_e^{\,2}} \frac{H_p^{(2)}(k_n^{\,e})}{H_p^{(2)'}(k_n^{\,e})} \right]. \quad (75)$$

Likewise, the expression (91) for the normalization factor N_n in the appendix can be somewhat simplified to give (92).

Instead of (72), the form (74) of the characteristic equation may be used; the wall impedance is in this case obtained from (73) by using the asymptotic value for the tangent function:

$$Z = \frac{\chi_n^{\,e}}{\omega \epsilon_e}. \quad (76)$$

4.3 Dielectric Matching Layer

The absorption of unwanted modes in the helix waveguide is critically dependent on the permittivity of the lossy jacket. In constructing a helix waveguide it is difficult to find a jacket material which has both a suitable value of complex permittivity and good enough mechanical properties to support the helix properly. A layer of lossless dielectric material in between helix and lossy jacket can serve to overcome both of these difficulties to a certain extent. By choosing a suitable material for this layer the bond between helix wires and jacket can be strengthened, thus improving the mechanical strength of the structure. Electrically, this layer acts as a radial transmission line-section. When made

of the proper thickness, this transmission line transforms the wall im-
pedance to more suitable values as far as unwanted mode attenuation
is concerned.

Let Z be the input impedance of the lossy jacket as given for example
by (73) or (76); then the new wall impedance in front of the dielectric
layer is given by the following transformation:

$$Z' = Z_1 \frac{Z + jZ_1 \tan \chi_{n1} l}{Z_1 + jZ \tan \chi_{n1} l}. \tag{77}$$

When ϵ_1 is the permittivity of the matching layer, the radial propaga-
tion constant is

$$\chi_{n1} = \sqrt{\omega^2 \mu \epsilon_1 - h_n{}^2}$$

and the wave impedance is

$$Z_1 = \frac{\chi_{n1}}{\omega \epsilon_1}.$$

The expression in (77) must be substituted for the wall impedance in
the characteristic equation (74).

4.4 *Laminated Jacket*

The foregoing consideration of one dielectric layer can be extended to
a semi-infinite stack of laminated material having alternate thin layers
of lossy and nonlossy material. It has indeed been proved very useful
to have the helix jacket made of such laminations. Excellent mechanical
and electrical properties have been achieved in this structure.

The wall impedance presented by a semi-infinite stack of alternating
layers is found by iterating the single layer transformation of (77). This
calculation has been made in a theory of laminated transmission lines:[8]

$$Z = \frac{T_{11} - T_{22}}{2T_{21}} + \frac{1}{2T_{21}} \sqrt{(T_{11} + T_{22})^2 - 4}. \tag{78}$$
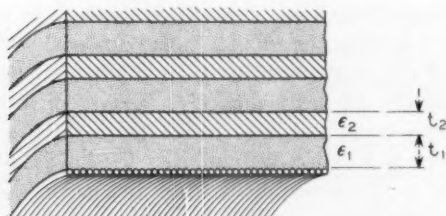
Fig. 2 — Laminated jacket of helix waveguide.

The T_{ik} are the elements of a transmission matrix relating input values of the electric and magnetic field strengths to the output values in one elementary double layer of the stack. With the constants and dimensions of Fig. 2, they are defined by

$$T_{11} = \cos \chi_{n1}l_1 \cos \chi_{n2}l_2 - \frac{Z_1}{Z_2} \sin \chi_{n1}l_1 \sin \chi_{n2}l_2 ,$$

$$T_{12} = jZ_1 \sin \chi_{n1}l_1 \cos \chi_{n2}l_2 + jZ_2 \sin \chi_{n2}l_2 \cos \chi_{n1}l_1 ,$$

$$T_{21} = \frac{j}{Z_1} \sin \chi_{n1}l_1 \cos \chi_{n2}l_2 + \frac{j}{Z_2} \sin \chi_{n2}l_2 \cos \chi_{n1}l_1 , \tag{79}$$

$$T_{22} = \cos \chi_{n1}l_1 \cos \chi_{n2}l_2 - \frac{Z_2}{Z_1} \sin \chi_{n1}l_1 \sin \chi_{n2}l_2 ,$$

where

$$\chi_{n1}^2 = \omega^2 \mu \epsilon_1 - h_n^2 ,$$

$$\chi_{n2}^2 = \omega^2 \mu \epsilon_2 - h_n^2$$

are the radial propagation constants in each individual layer and

$$Z_1 = \frac{\chi_{n1}}{\omega \epsilon_1} ,$$

$$Z_2 = \frac{\chi_{n2}}{\omega \epsilon_2} ,$$

are the radial wave impedances in each individual layer.

If the laminates are fine enough

$$| \chi_{n1}l_1 | \ll 1,$$

$$| \chi_{n2}l_2 | \ll 1,$$

the trigonometric functions may be replaced by their Taylor series. Then the wall impedance of (78) reduces to

$$Z = \frac{\chi_{nr}}{\omega \sqrt{\epsilon_z \epsilon_r}} - j \frac{1}{2} \frac{l_1 l_2}{\omega \epsilon_1 \epsilon_2} \frac{\omega^2 \mu \epsilon_1 \epsilon_2 - h_n^2 (\epsilon_1 + \epsilon_2)}{\epsilon_2 l_2 + \epsilon_1 l_1} (\epsilon_1 - \epsilon_2). \tag{80}$$

The first term of this expression is the radial wave impedance of a homogeneous but anisotropic medium. The permittivity of this medium is, for an electric field polarized in radial direction,

$$\epsilon_r = \frac{\epsilon_1 \epsilon_2 (l_1 + l_1)}{\epsilon_2 l_1 + \epsilon_1 l_2} \tag{81}$$

and, for an electric field polarized in longitudinal or azimuthal direction,

$$\epsilon_z = \frac{\epsilon_1 t_1 + \epsilon_2 t_2}{t_1 + t_2}. \tag{82}$$

The propagation constant of this anisotropic medium in (80) is

$$\chi_{nr}^2 = \omega^2 \mu \epsilon_r - h_n^2. \tag{83}$$

The second term in (80) is a correction for the finite thickness of the laminae.

For calculating wave propagation in a helix waveguide with a laminated jacket, (78) or (80) must be substituted for the wall impedance in the characteristic equation (74).

4.5 *Finite Size of Helix Wires*

In the preceding analysis the wires of the closely wound helix had been assumed to be of so small a size that the helix could be represented by an infinitely thin sheet ideally conducting in azimuthal direction and nonconducting in longitudinal direction. This anisotropic conducting sheath, however, does not always represent the helix satisfactorily. Helices wound from the smallest feasible wire sizes — 3 to 10 mils diameter (American Wire Gauge Nos. 30 to 40) — are not completely permeable to longitudinal electric fields. Displacement currents partly bridge the gap between successive turns. By partly shielding the jacket from the helix interior, they may change the propagation characteristics substantially.

This shielding effect is most easily taken into account by a capacitance in parallel to the input impedance of the jacket. The capacitance of an infinite grating of cylindrical wires is calculated in the Appendix. With the dimensions of Fig. 3, the capacitance per square of the helix surface is given by

$$C/\text{square} = \epsilon d \left(\frac{d}{D - d} - \frac{\ln 4}{\pi} \right),$$
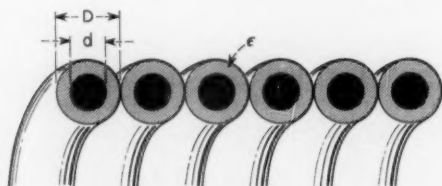
Fig. 3 — Helix wires.

where ϵ is the permittivity of the space in between wires. The corrected wall impedance is then obtained from

$$\frac{1}{Z'} = \frac{1}{Z} + j\omega\epsilon d \left(\frac{d}{D-d} - \frac{\ln 4}{\pi} \right). \tag{84}$$

Again, this expression must be substituted for the wall impedance in the characteristic equation (74).

All expressions for the wall impedance [(73), (76), (77), (78), (80), (84)] contain implicitly the axial propagation constant jh_n of the particular mode under consideration, and it is only through the characteristic equation (74) that this propagation constant will be determined.

In Section II, on the other hand, where the helix waveguide propagation is expressed in terms of normal modes of the solid wall guide, the wall impedance as introduced in the boundary condition (18) is not associated with a particular mode. The propagation constant jh_n remains undetermined. The boundary condition (18) cannot exactly replace a complex jacket structure.

For all cases, however, in which the modes under consideration are far from cutoff, the propagation constants are nearly equal to that of free space. We may then replace h_n by $\omega\sqrt{\mu_0\epsilon_0}$ in all expressions for the wall impedance and use them in (18) and the corresponding equations.

V. APPLICATIONS

Three different applications of the helix waveguide for circular electric wave transmission have been mentioned in the introduction. The formulae of the preceding sections will now be used to calculate the transmission properties of helix waveguide in the different applications. The theoretical values will be compared with results of measurements.

5.1 *Helix Waveguide Mode Filter*

It has been shown both theoretically and experimentally that the transmission characteristic of the circular waveguide can be substantially improved by the insertion of mode filters at intervals along the waveguide.[9] Conversion of energy to unwanted modes and reconversion at any imperfections of the waveguide seriously disturb the TE_{01} transmission characteristic. Mode filters, which provide low loss for the TE_{01} mode and high loss for all unwanted modes, greatly reduce the effects of mode conversion-reconversion by dissipating most of the converted power in unwanted modes before reconversion.

Helix waveguide has the desired properties of low TE_{0n} attenuation and high loss for all other modes. It therefore is well suited for a mode

filter. To design a helix waveguide mode filter one would choose a lossy material for the jacket which maximizes the attenuation constants of the unwanted modes of propagation. A detailed analysis[2] of the helix waveguide modes has shown, however, that there is no unique solution to this problem. As one changes the complex permittivity of the jacket, the attenuation constants of some modes increase while those of others decrease after passing through a maximum. Furthermore, a normal mode of a metallic waveguide incident in a helix waveguide will scatter its power into quite a few modes of the helix waveguide. As a result, a helix waveguide section inserted into a metallic waveguide will not only dissipate power in unwanted modes but will also introduce cross-coupling between unwanted modes. To find the transmission and conversion properties of a helix waveguide in the arrangement of Fig. 4 the matrix L of (42), as a scattering matrix between metallic waveguide modes and helix waveguide modes, must be combined with the transmission matrix T of the helix waveguide to give the relation between mode amplitudes at the input A_i and output A_0 :

$$A_0 = LTL_t A_i , \tag{85}$$

where T is a diagonal matrix, the elements of which are transmission coefficients of the helix waveguide modes

$$t_{nn} = e^{-jh_n z}. \tag{86}$$

Equation (85) has been evaluated for a typical helix waveguide with a relative permittivity $\epsilon/\epsilon_0 = 4 - j1$, and an inner radius to wavelength ratio $a/\lambda = 4.70$. This evaluation was extended over the TE_{11} , TM_{11} and TE_{12} modes of the metallic waveguide, and it was sufficient to take only five modes (TE_{11} , TE_{12} , TM_{12} , TE_{13} and TM_{13}) of the helix waveguide into account. The propagation constants jh_n of these helix waveguide modes were taken from the previously cited analysis of the helix waveguide.[2] For the elements of L, expressions (71) were substituted. The resulting transmission loss and conversion loss curves are shown in Fig. 5(a).
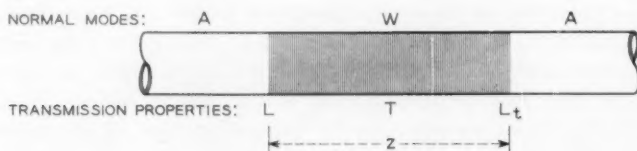
Fig. 4 — Matrices of mode filter.
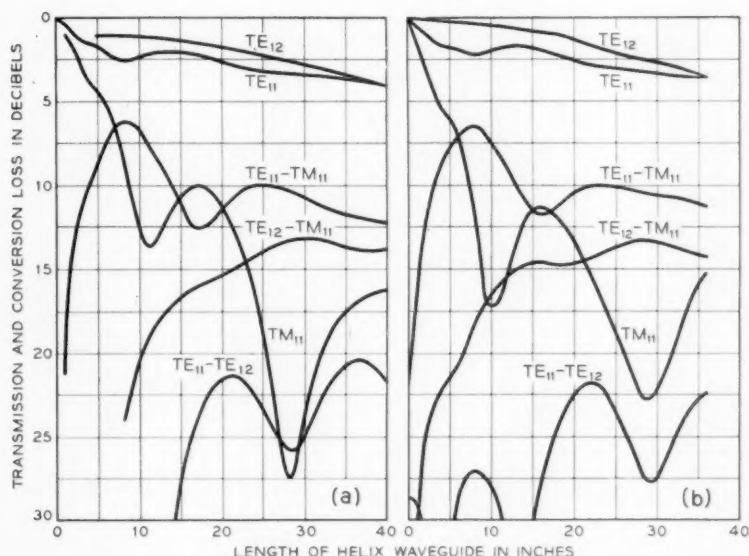
Fig. 5 — (a) Calculated transmission and conversion loss in helix waveguide: $a/\lambda = 4.70$, semi-infinite jacket with $\epsilon/\epsilon_0 = 4 - j1$; (b) measured transmission and conversion loss in helix waveguide of Fig. 6, $f = 55.5$ kmc.

The results of measurements on an 18 in. long section of 2 in. I.D. helix waveguide are shown in Fig. 5(b). This measured waveguide section has the helix wire dimensions and a jacket structure as shown in Fig. 6. The values of anisotropic permittivity of the laminated lossy material have been determined separately by impedance measurements on small samples in RG-98/U waveguide. The radial impedance of this laminated material is given by the first term in (80). The layer of glass roving between the lossy laminate and the helix wires transforms this impedance according to (77). Finally, the capacitance between helix wires changes the radial impedance according to (84). The resulting wall impedance is $Z = (147 + j22)\Omega$. The mathematical model of a helix waveguide which was used to calculate the curves of Fig. 5(a) has an isotropic and homogeneous jacket of relative permittivity $\epsilon/\epsilon_0 = 4 - j1$. Its wall impedance as calculated from (76) is $Z = (162 + j14)\Omega$. Comparing both wall impedances we find that the measured helix waveguide of Figs. 5(b) and 6 and the mathematical model of Fig. 5(a) are nearly equivalent. Thus the calculated curves of Fig. 5(a) may be compared
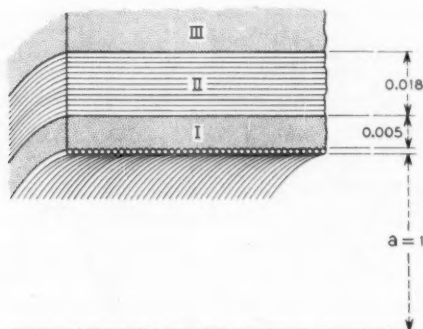
Fig. 6 — Helix waveguide structure. Helix: No. 37 copper wire with heavy Formex coat; Mediums I and III: Fiberglas laminated with epoxy resin; Medium II: six layers of tin oxide-coated glass cloth; $\epsilon_z/\epsilon_0 = 9 - j13$, $\epsilon_r/\epsilon_0 = 5 - j1$ at 55.5 kmc.

to the measured curves of Fig. 5(b). We find indeed very good agreement.

The transmission and conversion loss of the various modes were measured with millimeter-wave short pulse test equipment,[10] using the arrangement shown in Fig. 7. The pulses had a base width of about 3 millimicroseconds. Mode discrimination in the mode couplers and mode transducers combined with time discrimination based on the different group velocities of modes in the 220 ft. long waveguide provided sufficient over-all discrimination to identify energy travelling in different modes.

The results of Fig. 5 do not give an answer for designing helix waveguide for mode filtering purposes. They do, however, give assurance that the theory gives correct results even for quite complicated jacket structures, and that this theory may safely be used to design helix waveguide mode filters.
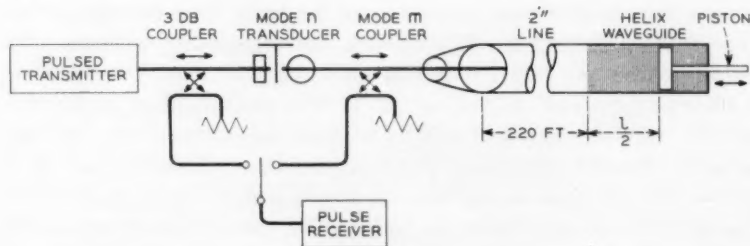
Fig. 7 — Measurement of transmission and conversion loss in helix waveguide.

The high amount of mode conversion which is caused by a helix waveguide of a certain length is perhaps surprising. The mode conversion loss between TE_{11} and TM_{11} is as low as 6 db in the case treated above. There may be cases in which this high mode conversion is undesirable. In order to reduce it, the abrupt transition from solid waveguide to helix waveguide may be replaced by a taper in wall impedance. The formulae of Section II give the basis to analyze such a wall impedance taper. The generalized telegraphist's equations in this section must then be solved for varying coupling coefficients instead of constant coupling coefficients.

5.2 Intentional Bends in Helix Waveguide

The analysis of a tapered curvature bend[3] has shown that the circular electric wave can be transmitted around bends with very low losses in a waveguide in which the degeneracy of equal phase velocity between TE_{01} and TM_{11} is broken up by lossless means. The shielded helix waveguide having a low-loss dielectric between the helix and the highly conducting metallic shield lends itself perfectly to this application.

A curvature taper on both sides of the bend serves to transform the normal mode of the straight pipe into the normal mode of the curved region and *vice versa*. A practical form for a tapered curvature bend is shown in Fig. 8; the region of constant curvature has vanished and all that remains is a triangular curvature distribution. This curvature dis-
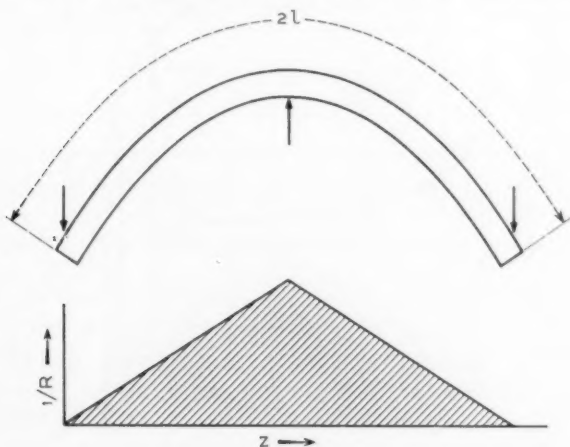
Fig. 8 — Triangular curvature distribution of an elastic deflection.

tribution can easily be made by bending the waveguide elastically over a fixed center support with forces acting on both ends.

The residual mode conversion in such a triangular curvature distribution is calculated by solving the coupled line equations for linearly varying coupling coefficients with suitable approximations for gentle curvature. Thus, at the end of a linear curvature taper of length l each mode n coupled to TE$_{01}$ through curvature causes a mode conversion loss of

$$A_{cn} = \operatorname{Re}\left[\frac{c_{0n}^{2}}{\Delta\gamma_{n}^{4}l^{2}}\left(1 - e^{-\Delta\gamma_{n}l} - \Delta\gamma_{n}l\right)\right] \text{ nepers}, \qquad (87)$$

where c_{0n} is the coefficient of curvature coupling (66) between TE$_{01}$ and a coupled mode n at the point of maximum curvature, and $\Delta\gamma_{n}$ is the difference in propagation-constant between the coupled mode n and TE$_{01}$.
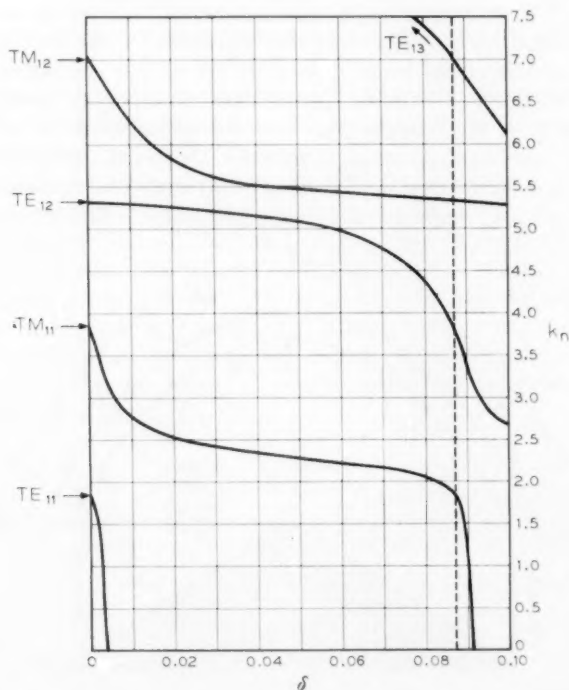
Fig. 9 — Solution of characteristic equation for shielded helix waveguide with lossless jacket; $\epsilon/\epsilon_0 = 2.5$, $a/\lambda = 4.70$.
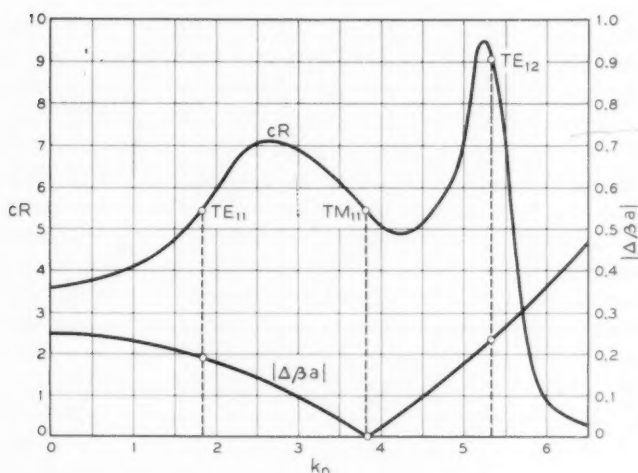
Fig. 10 — Difference in phase constant and coefficient of curvature coupling to TE_{01} in shielded helix waveguide; $a/\lambda = 4.70$.

The TE_{01} mode of the straight waveguide in passing through the curvature taper is transformed to the local normal mode of the curved region. It suffers a slight change in field configuration and consequently a change in attenuation constant. Thus, in a linear curvature taper of length l each mode coupled to TE_{01} by curvature causes an additional normal mode loss of

$$A_{bn} = \operatorname{Re}\left[\frac{c_{0n}^2 l}{3\Delta\gamma_n}\right] \text{nepers.} \qquad (88)$$

There is another way to calculate this normal mode loss: D. Marcuse has solved the boundary problem of the curved helix waveguide directly for the TE_{01} wave.[11] He obtains a closed expression for the curvature attenuation, which is in most cases easier to evaluate than a sum of the terms given by (88).

To provide data for the design of a low-loss helix waveguide for intentional bends, several of the expressions in Section III have been evaluated. In Fig. 9 the solution of the characteristic equation (72) is plotted in the range of interest for $p = 1$. A relative permittivity of 2.5 is a typical value for low loss dielectric materials. The ratio $a/\lambda = 4.70$ corresponds to a 2 in. I.D. helix waveguide operated at 55.5 kmc. In Fig. 10 the coefficient of curvature coupling from (66) and the difference in phase constant have been plotted for this helix waveguide. The mode
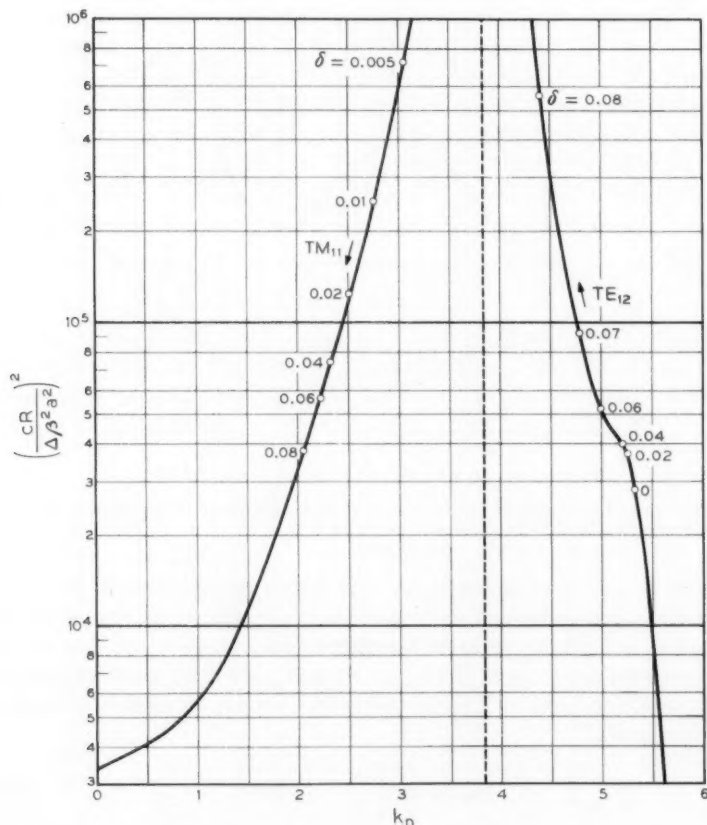
Fig. 11 — Factor of mode conversion loss in linear curvature taper of shielded helix waveguide; $\epsilon/\epsilon_0 = 2.5$, $a/\lambda = 4.70$.

conversion factor shown in Fig. 11 is found by combining the two curves of Fig. 10. This factor facilitates the calculation of mode conversion in the tapered curvature bend of Fig. 8.

An inspection of these conversion loss figures as well as of D. Marcuse's normal mode loss figures shows that, for optimum performance in a tapered curvature bend, the distance between helix and shield should be about a quarter of the radial wavelength in the jacket material. In no case should this distance approach a half radial wavelength anywhere in the frequency range.

Shielded helix waveguide with a low-loss jacket has been made with materials and construction techniques presently available in Bell Telephone Laboratories. The details of this waveguide are shown in Fig. 12. The TE_{01} loss in this structure was measured by a resonant cavity method both for the straight guide and the guide bent to various curves between 0 and 4.2 degrees. The curvature distribution was that of Fig. 8. Reasonable agreement between measured and calculated values of curvature loss was found. The measured loss followed closely the square-law dependence on curvature. At a total bending angle of 4.2 degrees the additional curvature loss in the 9 ft. long curved section was measured at 55.5 kmc to be 0.004 db. This figure is nearly equivalent to doubling the length of the waveguide. Mode filters on both sides of the curved section presented enough dissipation to avoid spurious mode interference in the measurements.

5.3 *All-Helix Waveguide*

A third and probably the most important application is a transmission line entirely consisting of helix waveguide. Present experimental experience shows that an all-helix waveguide line must present enough loss to unwanted modes so that power in these modes is dissipated at the same rate as it is converted at imperfections of the waveguide. In addition, the jacket of the helix waveguide should be designed so as to minimize the circular electric wave losses in bends, allowing relaxation of the straightness tolerances. Obviously, these two requirements contradict each other to a certain extent. A compromise will have to be chosen
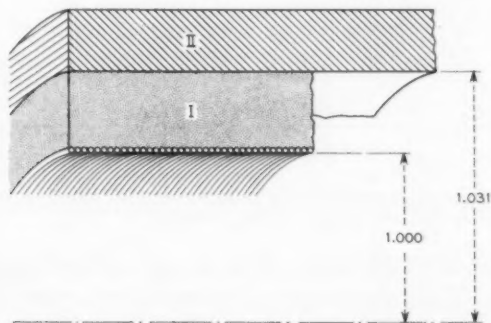
Fig. 12 — Shielded helix waveguide with low loss jacket. Helix: No. 37 copper wire with heavy Formex coat; Medium I: glass cloth laminated with epoxy resin, $\epsilon/\epsilon_0 = 4 - j0.1$ at 55.5 kmc; Medium II: copper tubing.
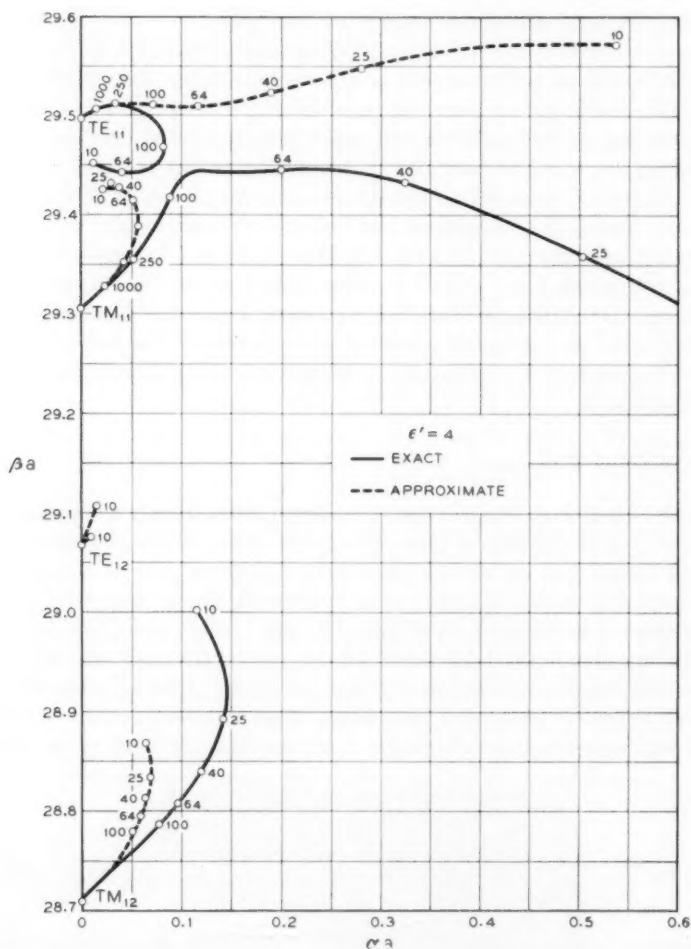
Fig. 13 — Propagation constants in helix waveguide; $a/\lambda = 4.70$, $\epsilon' = 4$. Representative values of ϵ'' are shown on the curves.

between sufficiently high unwanted mode loss and sufficiently low TE$_{01}$ curvature loss. This compromise will be made on the basis of tolerances. While the manufacturing tolerances call for a mode-filtering helix waveguide, the laying tolerances call for low TE$_{01}$ loss in random curvature.

However, it is to be expected that, as the art of making helix waveguide is improved and imperfections can be controlled better, the re-

quirement of low curvature loss will become more important and less
unwanted mode loss can be tolerated.

The first information needed for the design of a helix waveguide in
this application is the unwanted mode loss. It is contained in a detailed
analysis of the unwanted mode propagation in helix waveguide.[2] Some
of the results are taken from this analysis and plotted in Figs. 13 and 14.

Also plotted in Figs. 13 and 14 are the approximate values of the
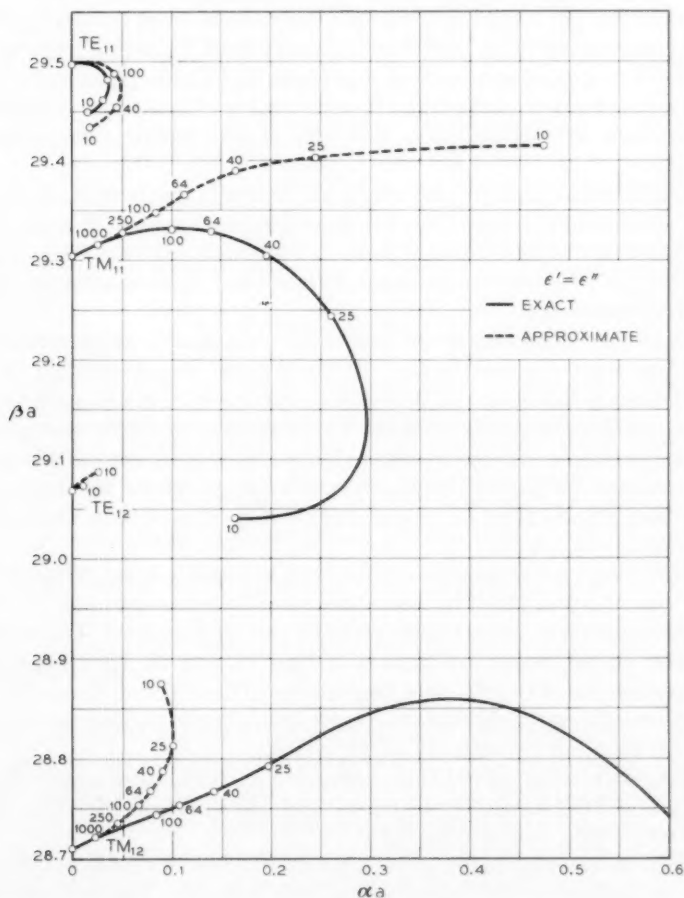
Fig. 14 — Propagation constants in helix waveguide; $a/\lambda = 4.70$, $\epsilon' = \epsilon''$. Representative values of ϵ'' are shown on the curves.

propagation constants which have been obtained as solutions of the characteristic equation (44) in Section II. The characteristic equation (44) has been reduced to a quartic equation by taking only four modes of the metallic guide (TE_{11}, TM_{11}, TE_{12}, TM_{12}) into account to represent the lower order modes with $p = 1$ in the helix waveguide. From Figs. 13 and 14 it has to be concluded that this representation is obviously a very poor one. At still fairly high values of the jacket permittivity, there is already a marked deviation of the approximate values from the exact curves. For a better representation in this large waveguide size, many more modes of the metallic guide have to be taken into account. It is therefore more convenient in most helix waveguide problems to use the representation of Section III with modes of the helix waveguide rather than the representation of Section II with modes of the metallic guide.

Imperfections like tilts and offsets or changes in diameter in an otherwise ideal guide convert TE_{01} power to unwanted modes. The effect of most of these imperfections in metallic waveguide has been analyzed in detail by S. P. Morgan, Jr., as quoted elsewhere.[1] More recently, another analysis has been made.[12]

For the helix waveguide the results of this analysis can be combined with the scattering matrix L of (42) to give the conversion from TE_{01} to the unwanted modes in the helix waveguide at these imperfections. Thus, given a certain distribution of imperfections in a helix waveguide, mode conversion and reconversion effects can be calculated. From such calculations the required unwanted mode loss to reduce the degrading effects of imperfections on TE_{01} propagation to an acceptable degree can be found.

The second requirement for an all-helix waveguide is low TE_{01} curvature loss. To find the TE_{01} curvature loss the expressions (47) or (66) for the coefficient of curvature coupling can be evaluated. The results of such an evaluation are plotted in Figs. 15 and 16. Here again the approximation (47) with only four modes (TE_{11}, TM_{11}, TE_{12}, TM_{12}) of the metallic guide deviates markedly from the exact values according to (66) for a low permittivity of the jacket.

Note that in the curved helix waveguide the TE_{01}-wave couples to all TM_{1n} and TE_{1n} waves. In the curved metallic waveguide the TE_{01} couples to all TE_{1n} but only to TM_{11}.

With the coupling coefficients on hand, TE_{01} loss and mode conversion can be calculated with formulae like (49) and (50). The perturbation of the TE_{01} propagation constant in (49), especially, means an increase α_c of TE_{01} attenuation. In Fig. 17 this attenuation increase α_c has been
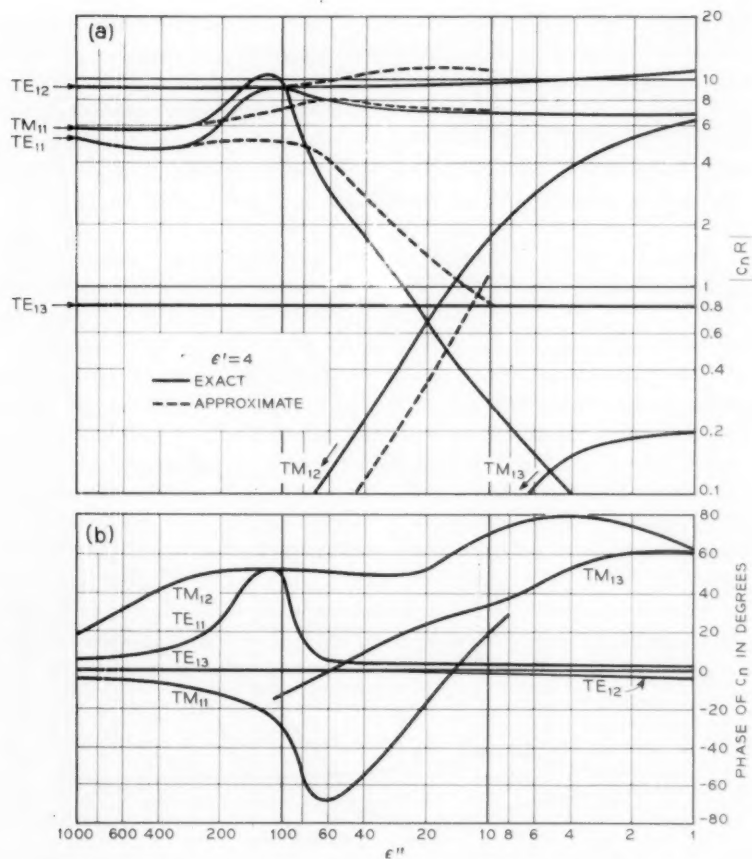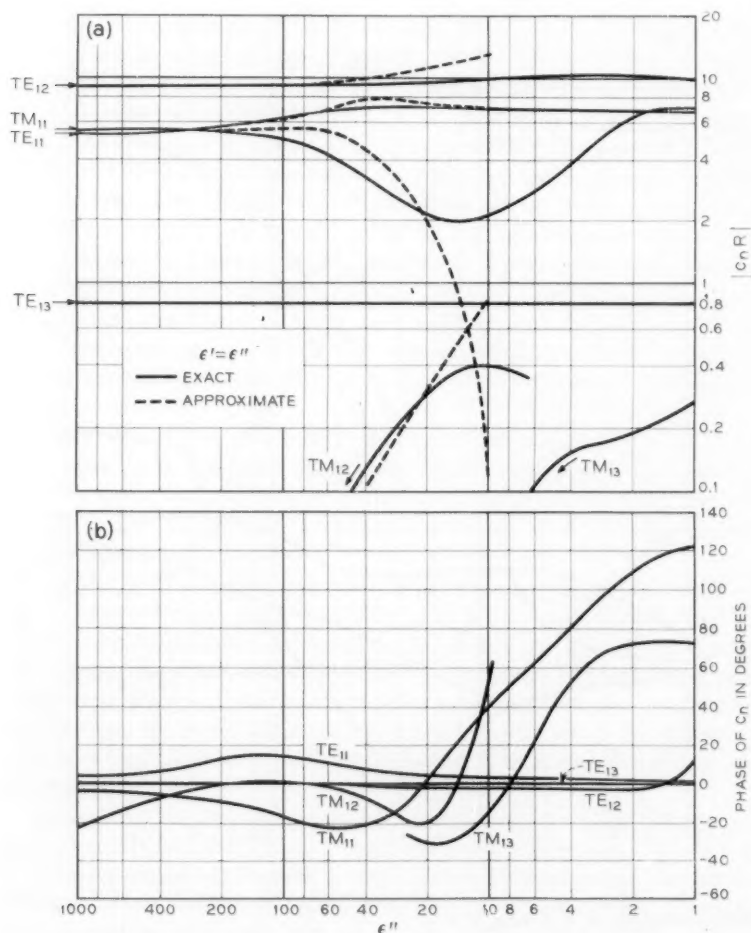
Fig. 15 — Coupling coefficient between TE_{01} and unwanted modes in curved helix waveguide; $a/\lambda = 4.70$, $\epsilon' = 4$.

plotted in a reduced form. The approximate values have again been obtained from the representation of Section II taking only four modes (TE_{11}, TM_{11}, TE_{12}, TM_{12}) into account. They deviate substantially from the exact values as obtained from Section III and equation (66). As mentioned earlier, α_c can also and more conveniently be calculated from D. Marcuse's normal mode solution of the curved helix waveguide.[11] Values for α_c of that calculation coincide with the exact curves of Fig. 17.

Fig. 16 — Coupling coefficient between TE_{01} and unwanted modes in curved helix waveguide; $a/\lambda = 4.70$, $\epsilon' = \epsilon''$.

The mode conversion which occurs in regions of changing curvature can be calculated from (50) if the change in curvature is abrupt, or from a solution of the coupled line equations with changing coupling coefficients when the curvature is changing gradually. Because of the laws of elasticity, the most common kind of curvature change, either intentionally or unintentional, is the linear curvature taper. The mode conversion

loss in a linear curvature taper is given by (87). From this expression it can be estimated if mode conversion effects in a changing curvature are large enough to influence the design of a helix waveguide. Usually it is found that this mode conversion is small enough to be neglected and that only the curvature attenuation α_c must be made as small as possible to optimize the performance of an all-helix waveguide in bends.

Here again, as in the case of the shielded helix waveguide with a loss-less jacket, a certain distance between helix wires and shield minimizes the curvature attenuation. The distance should be about a quarter of the radial wavelength in the jacket material. It should approach half of this wavelength nowhere in the frequency range.

Measurements of TE_{01} attenuation in bends have been made in the two different helix waveguides shown in Fig. 18. Eight-foot-long sections of the waveguides were bent elastically to the triangular curvature distribution of Fig. 8. The average curvature of this distribution is found by taking the square-law dependence between loss and curvature into account. The TE_{01} transmission loss was measured with a resonant cavity method for varying degrees of curvature. In the lossless helix waveguide, for intentional bends the TE_{01} transmission loss is composed of normal mode attenuation and mode conversion loss. However, in the present test of the lossy helix waveguide mode conversion losses can be neglected as compared to normal mode attenuation. The measured in-
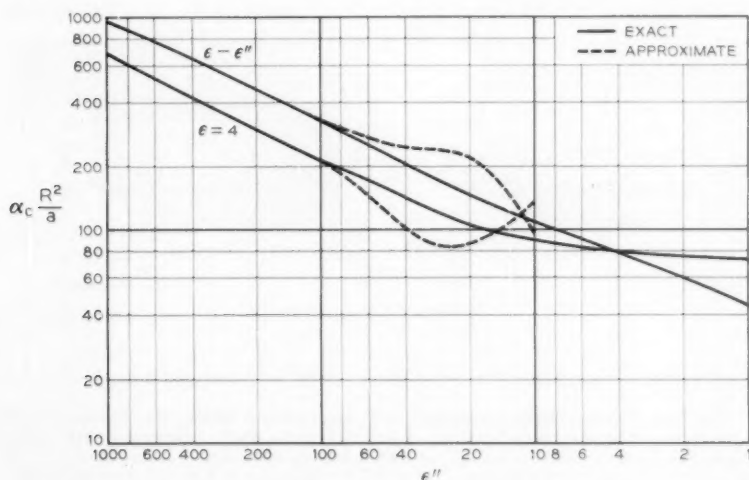
Fig. 17 — Increase of TE_{01} attenuation in curved helix waveguide; $a/\lambda = 4.70$.

crease in TE_{01} transmission loss in the curved waveguide can therefore be interpreted as additional TE_{01} attenuation per unit length. The results follow closely the square-law dependence on curvature and agree with theoretical values. The lower value of curvature attenuation in the second helix waveguide doubles the theoretical TE_{01} attenuation at a radius of curvature of 320 ft.

VI. CONCLUSIONS

Circular electric wave loss, mode conversion and unwanted mode loss can be calculated from generalized telegraphist's equations for the curved helix waveguide. A wall impedance can be used to represent the boundary condition at the helix interface. The effect of a composite jacket structure and of finite size helix wires can be taken into account in this wall impedance.

Three different applications of shielded helix waveguide for circular electric wave transmission require different designs for optimum performance. Helix waveguide as mode filter should have a highly lossy jacket between helix and shield for high unwanted mode absorption. An all-helix waveguide transmission line should have a medium lossy jacket to reduce TE_{01} loss in gentle bends and still maintain sufficient unwanted
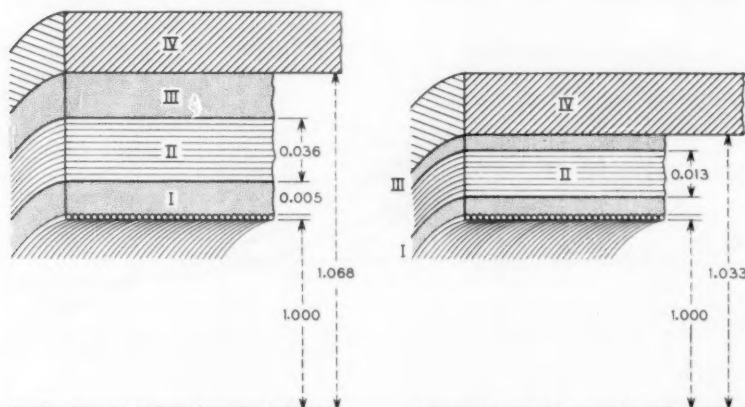
Fig. 18 — Shielded helix waveguide with lossy jacket. Helix: No. 37 copper wire with heavy Formex coat; Mediums I and III: glass cloth laminated with epoxy resin; Medium II: carbon loaded paper, laminated with epoxy resin, (left) $\epsilon_z/\epsilon_0 = 15 - j14$, $\epsilon_r/\epsilon_0 = 9 - j2$; (right) $\epsilon_z/\epsilon_0 = 4.7 - j2.1$, $\epsilon_r/\epsilon_0 = 3.9 - 0.6$ at 55.5 kmc; Medium IV: steel tubing. TE_{01} curvature loss $\alpha_c(R^2/a)$ at 55.5 kmc: (left) measured, 79, calculated, 84; (right) measured, 48, calculated, 40.

mode absorption. For sharp intentional bends, the material between helix and shield should have low losses and the shield should be highly conducting. In all three applications the optimum distance between helix and shield is a quarter of the radial wavelength in the jacket material.

APPENDIX

Normalization of Modes in Helix Waveguide

The orthogonality relation (60) for $n \neq m$ has been proven for exponential modes on an inhomogeneous cylindrical structure bounded by opaque walls or unbounded.[13]

The right-hand side of (60) can be rewritten as

$$\frac{1}{V_n I_m} \int_s (E_{tn} \times H_{tm}) \, dS$$

$$= \int_s \frac{\epsilon}{\epsilon_0} \left[(\text{grad } T_n)(\text{grad } T_m) + d_n d_m (\text{grad } T_n{}')(\text{grad } T_m{}') \right. \quad (89)$$

$$\left. - d_m \frac{h_m{}^2}{k^2} (\text{flux } T_n)(\text{grad } T_m{}') + d_n (\text{flux } T_n{}')(\text{grad } T_m) \right] dS.$$

Partial integrations, like Green's theorem,

$$\int_s (\text{grad } T_n)(\text{grad } T_m) \, dS = \int_v T_m \frac{\partial T_n}{e_1 \partial u} e_2 \, dv + \int_s \chi_n{}^2 T_n T_m \, dS,$$

and

$$\int_s (\text{flux } T_n)(\text{grad } T_m) \, dS = - \int_v T_n \frac{\partial T_m}{e_2 \partial v} e_2 \, dv = \int_v T_m \frac{\partial T_n}{e_2 \partial v} e_2 \, dv$$

reduce (89) to:

$$\int_s \frac{\epsilon}{\epsilon_0} \chi_n^2 \left(T_n T_m + d_n d_m \frac{h_m^2}{k^2} T_n' T_m' \right) dS$$

$$+ \int_v \frac{\epsilon}{\epsilon_0} T_m \left[\frac{\partial T_n}{e_1 \partial u} + d_n \frac{\partial T_n'}{e_2 \partial v} \right] e_2 \, dv$$

$$- \int_v \frac{\epsilon}{\epsilon_0} d_m \frac{h_m^2}{k^2} T_m' \left[\frac{\partial T_n}{e_2 \partial v} - d_n \frac{\partial T_n'}{e_1 \partial u} \right] e_2 \, dv.$$

The integrand of the last term is continuous across the helix boundary and vanishes at the shield. It therefore cancels out;

$$\frac{1}{V_n I_m} \int (E_{tn} \times H_{tm}) \, dS = \int_s \frac{\epsilon}{\epsilon_0} \chi_n^2 \left(T_n T_m + d_n d_m \frac{h_m^2}{k^2} T_n' T_m' \right) dS \\ + \int_v \frac{\epsilon}{\epsilon_0} T_m \left(\frac{\partial T_n}{e_1 \partial u} + d_n \frac{\partial T_n'}{e_2 \partial v} \right) e_2 \, dv. \tag{90}$$

From (90) the orthogonality relation can be verified for $n \neq m$.

To calculate the normalization factor N_n we substitute (51) for the T functions in (90) and perform the integrations. Since in all cases of practical interest $|k_n^e| \gg |(4p^2 - 1)/8|$, the Hankel functions may be replaced by their asymptotic expressions. Then, in order to satisfy (60) the normalization factor has to be

$$N_n = \frac{\sqrt{2}}{\sqrt{\pi} \, J_p(k_n)} \left[\frac{h_n^2}{k^2} p^2 (k_n^2 - p^2) Y_n^2 + \frac{1}{Y_n^2} + k_n^2 \left(1 - \frac{p^2}{k^2 a^2} \right) \right. \\ \left. + 2 \left(\frac{1}{Y_n} - p^2 Y_n \right) + \frac{\epsilon_e}{\epsilon_0} \frac{k_n^4}{k_n^{e3}} \left(2 \cot \delta k_n^e + \frac{\delta k_n^e}{\sin^2 \delta k_n^e} \right) \right]^{-1/2}, \tag{91}$$

where

$$Y_n = \frac{J_p(k_n)}{k_n J_p'(k_n)}.$$

For the unshielded helix waveguide described in Section IV the normalization factor is given by:

$$N_n = \frac{\sqrt{2}}{\sqrt{\pi} \, J_p(k_n)} \left[\frac{h_n^2}{k^2} p^2 (k_n^2 - p^2) Y_n^2 + \frac{1}{Y_n^2} + k_n^2 \left(1 - \frac{p^2}{k^2 a^2} \right) \right. \\ \left. + 2 \left(\frac{1}{Y_n} - p^2 Y_n \right) + j2 \frac{\epsilon_e}{\epsilon_0} \frac{k_n^4}{k_n^{e3}} \right]^{-1/2}. \tag{92}$$

For the other jacket structures it is not an easy matter to calculate the normalization factor. Finding the field vectors in the jacket region

and integrating over the cross-product of the transverse components is quite involved, even when asymptotic approximations such as in (91) are substituted. A further approximation, however, eliminates this problem. The integral on the left-hand side of (89) is the flux of the Poynting vector, through a waveguide cross section. From the dimensions of the helix waveguide we may safely conclude that for a typical mode the power flowing in the jacket is small compared to the power flowing inside of the helix in all cases of practical interest. Then we may extend the integral in (89) only over the helix interior and neglect the last terms under the square root of (91) or (92).

Capacitance of Plane Grating of Cylindrical Wires

The contour in the z-plane of Fig. 19 is an elementary cell of the field surrounding a plane grating of cylindrical wires. Its transformation to the strip $0 \leqq u \leqq \pi$ of the w-plane is approximately effected by[14]

$$z = \frac{D}{\pi(1 + \nu)} \left[\tanh^{-1} \sqrt{\frac{z_1 - 1}{z_1 + a}} + \nu \tanh^{-1} \sqrt{\frac{z_1 + 1}{z_1 + a}} \right] \quad (93)$$

and

$$z_1 = \frac{a - 1}{2} \cos w - \frac{a + 1}{2}, \quad (94)$$
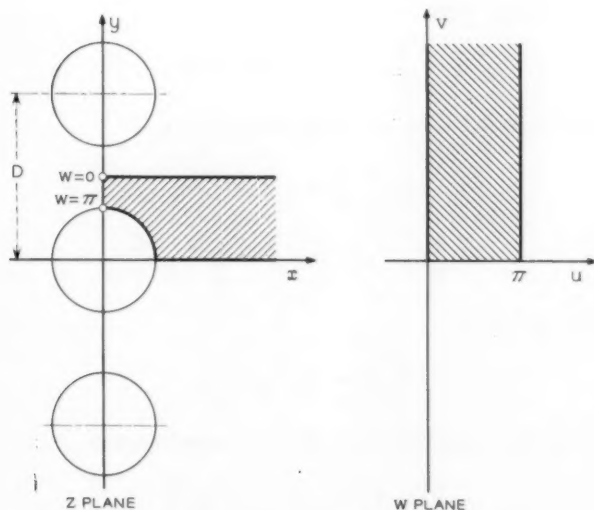
Fig. 19 — Conformal mapping of wire grating.

where ν is the root of

$$\coth^2\left(\frac{\pi}{c}\frac{d}{D}\frac{1+\nu}{\nu}\right) - \cot^2\frac{\pi}{2}\frac{d}{D}(1+\nu) = 1 \tag{95}$$

and a is given by

$$a = 1 + 2\cot^2\frac{\pi}{2}\frac{d}{D}(1+\nu). \tag{96}$$

A uniform electric field polarized in the direction of the y-axis is fringed by the grating. The capacitance of this fringing field is the difference between the capacitance of the z-plane contour and a parallel plate condenser of plate distance $D/2$ and extension $2x$:

$$\frac{C}{\epsilon} = \lim_{x\to\infty}\left(\frac{2}{\pi}v - \frac{4x}{D}\right). \tag{97}$$

Substituting for v from (93) and (94) we get, instead of (97),

$$\frac{C}{\epsilon} = \frac{2}{\pi}\left[\frac{\ln(a+1) + \nu\ln(a-1)}{1+\nu} - \ln(a-1)\right]. \tag{98}$$

For $D/d = 1$ the root of (95) is $\nu = 0$. When d is only slightly smaller than D, as is the case in a closely wound helix, ν is small and the argument of the cot in (95) is nearly $\pi/2$, while the argument of the coth is very large. Then, with

$$\frac{d}{D}(1+\nu) = (1+\eta)$$

the cot and coth functions can be approximated by:

$$\cot\frac{\pi}{2}\frac{d}{D}(1-\nu) = \frac{\pi}{2}\eta,$$

$$\coth\frac{\pi}{2}\frac{d}{D}\frac{(1+\nu)}{\nu} = 1 + 2e^{-\pi}\frac{d}{D-d}.$$

Instead of (95) we get

$$\frac{\pi^2}{4}\eta^2 = 4e^{-\pi}\frac{d}{D-d}.$$

Substituting for a from (96) into (98), the capacitance is

$$\frac{C}{\epsilon} = 2\frac{d}{D}\left(\frac{d}{D-d} - \frac{\ln 4}{\pi}\right). \tag{99}$$

REFERENCES

1. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, Nov. 1954, pp. 1209–1265.
2. Morgan, S. P. and Young, J. A., Helix Waveguide, B.S.T.J., **35**, Nov. 1956, pp. 1347–1384.
3. Unger, H. G., Normal Mode Bends for Circular Electric Waves, B.S.T.J., **36**, Sept. 1957, pp. 1292–1307.
4. Schelkunoff, S. A., Conversion of Maxwell's Equations into Generalized Telegraphist's Equations, B.S.T.J., **34**, Sept. 1955, pp. 995–1043.
5. Morgan, S. P., Theory of Curved Circular Waveguide Containing an Inhomogeneous Dielectric, B.S.T.J., **36**, Sept. 1957, pp. 1209–1251.
6. Guillemin, E. A., *The Mathematics of Circuit Analysis*, John Wiley & Sons, New York, 1950, pp. 111–122.
7. Rose, C. F. P., Research Models of Helix Waveguide, B.S.T.J., **37**, May 1958, pp. 679–688.
8. Morgan, S. P., Mathematical Theory of Laminated Transmission Lines, B.S.T.J., **31**, September 1952, pp. 883–949.
9. Warters, W. D., The Effects of Mode Filters on the Transmission Characteristics of Circular Electric Waves in a Circular Waveguide, B.S.T.J., **37**, May 1958, pp. 657–677.
10. Burrus, C. A., Millimicrosecond Pulses in the Millimeter Wave Region, Rev. Sci. Instr., **28**, Dec. 1957, pp. 1062–1065.
11. Marcuse, D., this issue, pp. 1649–1662.
12. Iiguchi, S., Mode Conversion in the Transmission of TE$_{01}$ Wave Through a Slight Tilt and a Slight Offset of Waveguide, J. Inst. Elec. Comm. Engrs. Japan, **40**, 1957, pp. 870–876; 1095–1102.
13. Bresler, A. D., Joshi, G. H. and Marcuvitz, N., Orthogonality Properties for Modes in Passive and Active Uniform Waveguides, J. Appl. Phys. **29**, pp. 794–799.
14. Smythe, W. R., *Static and Dynamic Electricity*, McGraw-Hill Book Co., New York, 1950, p. 98–100.
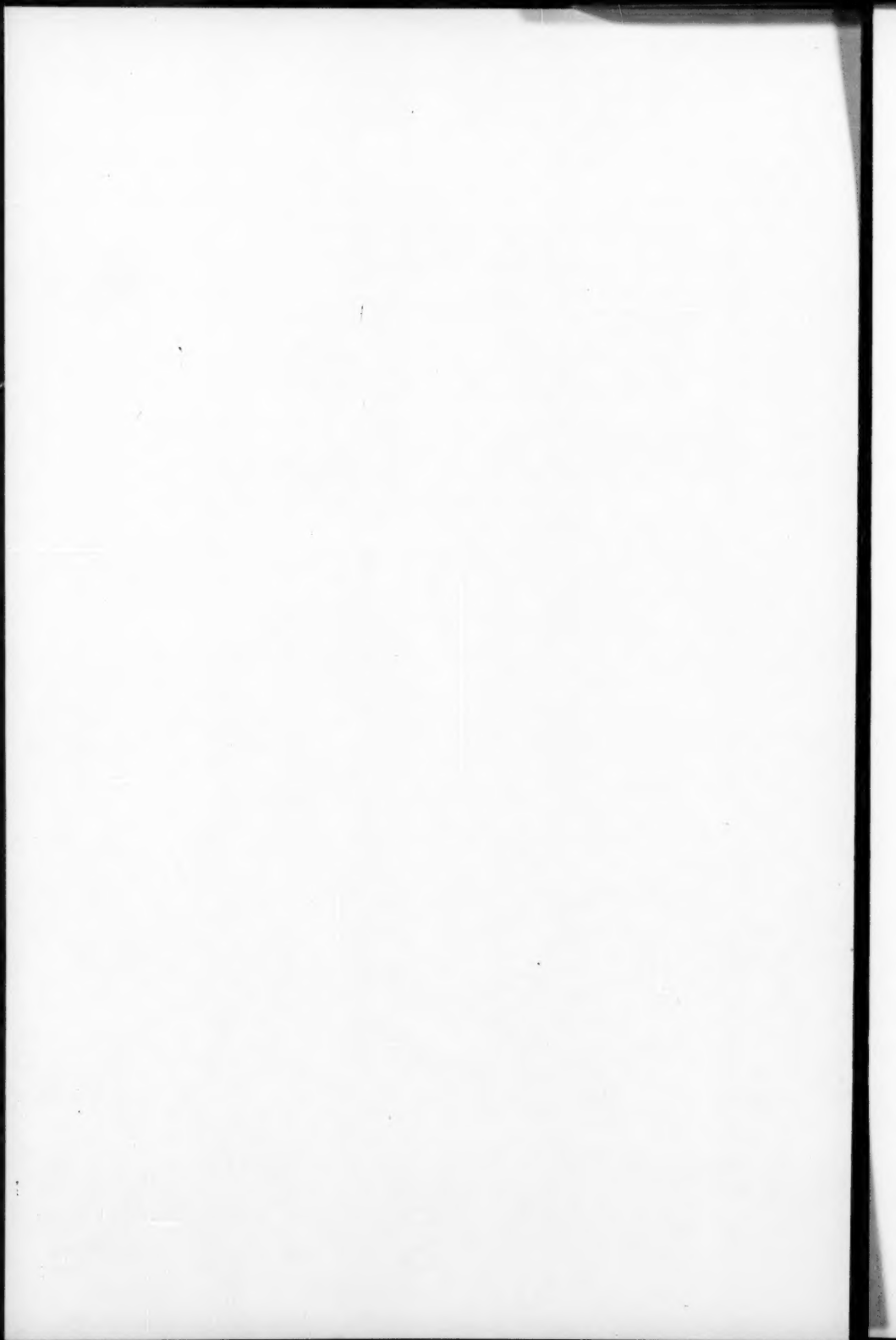
Attenuation of the TE_{01} Wave Within the Curved Helix Waveguide*

By D. MARCUSE

(Manuscript received April 1, 1958)

The change in the field pattern of the TE_{01} mode in the helix waveguide with constant curvature is calculated by means of a perturbation method. The helix waveguide has a coating of a lossy dielectric and is shielded by a metallic pipe. By bending this waveguide, the original field pattern of the TE_{01} wave is slightly perturbed. The perturbation produces additional field components which give rise to an electromagnetic field in the lossy dielectric. This field causes energy dissipation and gives rise to an additional attenuation. In this paper formulae for the curvature attenuation caused by the dielectric losses are given. It is shown that the attenuation due to curvature can be remarkably reduced by proper choice of the thickness of the dielectric layer between helix and metallic shield.

I. INTRODUCTION

If the system of cylindrical coordinates, in which the cylindrical waveguide ordinarily is described, is bent in the same way as the axis of the curved waveguide, Maxwell's equations change. Jouguet[1] has shown an approximate solution of Maxwell's equations written in a system of toroidal coordinates. This solution may be applied to calculate the field pattern of a TE_{01} wave in the curved helix waveguide.[2] The waveguide[3] consists of a helix made of very thin wire and a coating of a lossy dielectric which may be shielded by a metallic pipe. The lossy dielectric serves to suppress unwanted modes which may be excited in the helix waveguide. Although the lossy dielectric does not affect the TE_{01} wave in a perfectly straight guide, it does give rise to an additional attenuation if the waveguide is bent. It is the purpose of the present paper to show how a helix waveguide should be designed in order to keep this additional attenuation as low as possible. On the other hand, the attenuation of the TE_{01} wave in bends is not the only effect which is to be considered

---

* The author used the same method of analysis to calculate the attenuation and field structure of the TE_{01} wave in a curved spaced-ring type waveguide while he was working with the Siemens and Halske Company in Germany.

in designing the waveguide. In practice, a compromise should be made between the desire to keep the attenuation of the TE_{01} wave in a curved waveguide low and the desire to suppress spurious modes.[4] For a short part of the guide with a wanted high curvature, a helix waveguide which is designed for minimum curvature attenuation of the TE_{01} wave might be useful.

The mathematical model used for calculating the helix waveguide neglects the pitch angle of the helix wire.[5] It is assumed that the closely wound helix may be idealized as an infinitely thin sheet with anisotropic conductivity, conducting only in the direction perpendicular to the axis of the waveguide.

II. SOLUTION OF THE BOUNDARY VALUE PROBLEM

In the following calculation it is assumed that the helix and the pipe which serves as a shield around the helix waveguide are of perfect conductivity. The toroidal coordinates have the variables r, φ and z. These are the same variables as for an ordinary system of cylindrical coordinates. The only difference is that the z axis is curved in the same way as the axis of the helix waveguide. The line element is given as

$$ds^2 = \left(1 + \frac{r}{R}\sin\varphi\right)^2 dz^2 + dr^2 + r^2\, d\varphi^2,$$

where R is the radius of curvature which is assumed to be constant. Maxwell's equations are then given by:

$$\frac{\partial}{\partial\varphi}\left[\left(1 + \frac{r}{R}\sin\varphi\right)H_z\right] - \gamma(rH_\varphi) = j\omega\epsilon r\left(1 + \frac{r}{R}\sin\varphi\right)E_r,$$

$$\gamma H_r - \frac{\partial}{\partial r}\left[\left(1 + \frac{r}{R}\sin\varphi\right)H_z\right] = j\omega\epsilon\left(1 + \frac{r}{R}\sin\varphi\right)E_\varphi,$$

$$\frac{\partial}{\partial r}(rH_\varphi) - \frac{\partial}{\partial\varphi}H_r = j\omega\epsilon rE_z,$$

$$\frac{\partial}{\partial\varphi}\left[\left(1 + \frac{r}{R}\sin\varphi\right)E_z\right] - \gamma rE_\varphi = -j\omega\mu r\left(1 + \frac{r}{R}\sin\varphi\right)H_r,$$

$$\gamma E_r - \frac{\partial}{\partial r}\left[\left(1 + \frac{r}{R}\sin\varphi\right)E_z\right] = -j\omega\mu\left(1 + \frac{r}{R}\sin\varphi\right)H_\varphi,$$

$$\frac{\partial}{\partial r}(rE_\varphi) - \frac{\partial}{\partial\varphi}E_r = -j\omega\mu rH_z,$$

$$(1)$$

if the z and t dependence of all field components are of the form

$$e^{j\omega t + \gamma z}.$$

$$(2)$$

If R tends to ∞ the equations (1) reduce to the ordinary Maxwell's equations in cylindrical coordinates. Since $r/R \ll 1$ in all practical applications, the solutions of (1) will be nearly the same as the solutions for the system with straight z axis; they will differ only by additional perturbation terms. Therefore, the following statement is made:

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{h},$$
$$\mathbf{E} = \mathbf{E}_0 + \mathbf{e}, \tag{3}$$

where $\mathbf{H}_0$ and $\mathbf{E}_0$ are the solutions of (1) for $R \to \infty$. Since we are concerned with the special solutions which give the TE_{01} wave in the curved helix waveguide, the field of zero-order approximation will be the ordinary TE_{01} wave:

$$H_{z0} = AJ_0(\chi_0 r)e^{j\omega t + \gamma z},$$

$$H_{r0} = -\frac{\gamma}{\chi_0}AJ_1(\chi_0 r)e^{j\omega t + \gamma z},$$

$$E_{\varphi 0} = -j\frac{\omega\mu}{\chi_0}AJ_1(\chi_0 r)e^{j\omega t + \gamma z}, \tag{4}$$

$$\chi_0^2 = \beta_0^2 + \gamma^2,$$

$$\beta_0^2 = \omega^2 \epsilon_0 \mu_0.$$

Equations (4) and (3) are now substituted into (1). Since $\mathbf{h}$ and $\mathbf{e}$ must be of the same order of magnitude as r/R, products of these terms with r/R are omitted. We therefore get only the first-order approximation.

If we consider that $\mathbf{H}_0$ and $\mathbf{E}_0$ are solutions of the unperturbed system, (1) becomes

$$\frac{\partial}{\partial r}(rh_\varphi) - \frac{\partial}{\partial\varphi}h_r - j\omega\epsilon re_z = 0,$$

$$\frac{\partial}{\partial\varphi}h_z - \gamma rh_\varphi - j\omega\epsilon re_r = -\frac{r}{R}\cos\varphi AJ_0(\chi_0 r),$$

$$\gamma h_r - \frac{\partial}{\partial r}h_z - j\omega\epsilon e_\varphi = -\frac{r}{R}\frac{\gamma^2}{\chi_0}\sin\varphi AJ_1(\chi_0 r)$$
$$+ \frac{1}{R}\sin\varphi AJ_0(\chi_0 r), \tag{5}$$

$$\frac{\partial}{\partial r}(re_\varphi) - \frac{\partial}{\partial\varphi}e_r + j\omega\mu rh_z = 0,$$

$$\frac{\partial}{\partial\varphi}e_z - \gamma re_\varphi + j\omega\mu rh_r = \frac{r}{R}\gamma r\frac{j\omega\mu}{\chi_0}AJ_1(\chi_0 r),$$

$$\gamma e_r - \frac{\partial}{\partial r}e_z + j\omega\mu h_\varphi = 0.$$

For $R \to \infty$, these equations are the ordinary unperturbed Maxwell equations for $\mathbf{e}$ and $\mathbf{h}$. The terms with $1/R$ make the system of equations (5) inhomogeneous. Its solution is therefore given as the general solution of the homogeneous system and a particular solution of the inhomogeneous system. The solution of the inhomogeneous equations is found by the statement:

$$
\begin{aligned}
e_z &= X \cos \varphi, & h_z &= U \sin \varphi, \\
e_r &= Y \cos \varphi, & h_r &= V \sin \varphi, \\
e_\varphi &= \frac{Z}{r} \sin \varphi, & h_\varphi &= \frac{W}{r} \cos \varphi.
\end{aligned}
\tag{6}
$$

If we substitute this into (5) we get a system of ordinary differential equations with the only variable r. This system is solved by a statement of the form

$$
X = P J_0(\chi_0 r) + Q J_1(\chi_0 r),
$$

where P and Q are polynomials of r only. After determining the coefficients of the polynomials we finally get the solution of (5), after adding the solution of the homogeneous equation:

$$
h_z = \left[B J_1(\chi_0 r) - \frac{\chi_0^2 + 2\gamma^2}{2\chi_0^2} \, r \, \frac{A}{R} J_0(\chi_0 r) \right.
$$
$$
\left. + \frac{\gamma^2}{2\chi_0} r^2 \frac{A}{R} J_1(\chi_0 r) \right] e^{\gamma z} \sin \varphi,
$$

$$
h_r = \left[B \frac{\gamma}{\chi_0} J_1'(\chi_0 r) - jC \frac{\omega \epsilon_0}{\chi_0^2} \frac{1}{r} J_1(\chi_0 r) \right.
$$
$$
+ \frac{\gamma}{2\chi_0^2} (r^2 \gamma^2 - 1) \frac{A}{R} J_0(\chi_0 r)
$$
$$
\left. + \frac{\gamma}{2\chi_0^3} (\beta_0^2 + 2\chi_0^2) r \frac{A}{R} J_1(\chi_0 r) \right] e^{\gamma z} \sin \varphi,
$$

$$
h_\varphi = \left[B \frac{\gamma}{\chi_0^2} \frac{1}{r} J_1(\chi_0 r) - jC \frac{\omega \epsilon_0}{\chi_0} J_1'(\chi_0 r) - \frac{1}{2} \frac{\gamma}{\chi_0^2} \frac{A}{R} J_0(\chi_0 r) \right.
$$
$$
\left. + \frac{1}{2} \frac{\gamma}{\chi_0^3} (\beta_0^2 + \chi_0^2) r \frac{A}{R} J_1(\chi_0 r) \right] e^{\gamma z} \cos \varphi,
\tag{7}
$$

$$
e_z = \left[C J_1(\chi_0 r) - j\omega\mu \frac{\gamma}{\chi_0^2} r \frac{A}{R} J_0(\chi_0 r) \right] e^{\gamma z} \cos \varphi,
$$

$$e_r = \left[-jB\frac{\omega\mu}{\chi_0{}^2}\frac{1}{r}J_1(\chi_0 r) + C\frac{\gamma}{\chi_0}J_1{}'(\chi_0 r) - \frac{j\omega\mu}{2\chi_0{}^2}\frac{A}{R}J_0(\chi_0 r) \right.$$

$$\left. + \frac{j\omega\mu}{2\chi_0{}^3}r\gamma^2\frac{A}{R}J_1(\chi_0 r) \right]e^{\gamma z}\cos\varphi,$$

$$e_\varphi = \left[jB\frac{\omega\mu}{\chi_0}J_1{}'(\chi_0 r) - C\frac{\gamma}{\chi_0{}^2}\frac{1}{r}J_1(\chi_0 r) \right.$$

$$\left. + \frac{1}{2}\frac{j}{\omega\epsilon_0}\frac{\beta_0{}^2}{\chi_0{}^2}(1 + r^2\gamma^2)\frac{A}{R}J_0(\chi_0 r) + \frac{1}{2}j\omega\mu\frac{\beta_0{}^2}{\chi_0{}^3}r\frac{A}{R}J_1(\chi_0 r) \right]e^{\gamma z}\sin\varphi.$$

The validity of these solutions can be proven by substituting into (5). The solution of the homogeneous equations is chosen so as to have the same φ dependence and the same propagation constant as the solution of the inhomogeneous equation.

Up to this point the theory has followed the solution given in Jouguet's paper and is not restricted to the application to the helix waveguide.

Equations (7) and (4) give the field pattern of a normal mode in the curved helix waveguide, and $\epsilon = \epsilon_0$ is taken for a guide filled with air. The field in the dielectric outside of the helix should also be a solution of (1). But since the TE$_{01}$ wave in a straight guide has no field outside of the helix, the field in the dielectric layer is excited only by the additional field components (7) and is therefore of the same order of magnitude. If we again neglect terms of the order $(r/R)^2$ we get Maxwell's equations for the straight waveguide for the small field within the dielectric layer. This field should have the same φ dependence as (7). We therefore state:

$$h_z{}^e = FW_1(\chi' r)\sin\varphi e^{\gamma z},$$

$$h_r{}^e = \left[F\frac{\gamma}{\chi'}W_1{}'(\chi' r) - jD\frac{\omega\epsilon}{\chi'^2}\frac{1}{r}V_1(\chi' r) \right]\sin\varphi e^{\gamma z},$$

$$h_\varphi{}^e = \left[F\frac{\gamma}{\chi'^2}\frac{1}{r}W_1(\chi' r) - jD\frac{\omega\epsilon}{\chi'}V_1{}'(\chi' r) \right]\cos\varphi e^{\gamma z},$$

$$e_z{}^e = DV_1(\chi' r)\cos\varphi,$$

$$e_r{}^e = \left[-jF\frac{\omega\mu_0}{\chi'^2}\frac{1}{r}W_1(\chi' r) + D\frac{\gamma}{\chi'}V_1{}'(\chi' r) \right]\cos\varphi e^{\gamma z}, \qquad (8)$$

$$e_\varphi{}^e = \left[jF\frac{\omega\mu_0}{\chi'}W_1{}'(\chi' r) - D\frac{\gamma}{\chi'^2}\frac{1}{r}V_1(\chi' r) \right]\sin\varphi e^{\gamma z},$$

$$\chi'^2 = \beta'^2 + \gamma^2,$$

$$\beta'^2 = \omega^2\epsilon\mu_0,$$

$$\frac{\epsilon}{\epsilon_0} = \epsilon' - j\epsilon'',$$

with $\epsilon' =$ dielectric constant, $\epsilon'' = \epsilon' \tan \delta$, where $\tan \delta$ is the loss factor of the dielectric material. The functions V and W are abbreviations for

$$V_1(\chi'r) = H_1^{(2)}(\chi'r) - \frac{H_1^{(2)}(\chi'b)}{H_1^{(1)}(\chi'b)} H_1^{(1)}(\chi'r),$$

$$W_1(\chi'r) = H_1^{(2)}(\chi'r) - \frac{H_1^{(2)'}(\chi'b)}{H_1^{(1)'}(\chi'b)} H_1^{(1)}(\chi'r),$$

(9)

where $H_1^{(1)}$ and $H_1^{(2)}$ are Hankel's functions of the first and second kind respectively. The prime on the V, W and Hankel function denotes differentiation with respect to the argument. The functions V and W are made such that e_z and e_φ vanish at the shield $r = b$. The boundary conditions on the shield are thus satisfied by the statements (8) and (9). The field of zero-order approximation has to be matched to the boundary independent of the additional field components because of their different dependence on the variable φ. It must be $E_{0\varphi} = 0$ at $r = a$. That means $J_1(\chi_0 a) = 0$, $\chi_0 a = 3.83$, which is exactly the same eigenvalue as that for the straight helix waveguide or even the ordinary circular solid-wall waveguide. To first order of approximation there is no change in eigenvalue. The field components (7) and (8) must be matched at the helix $r = a$ by means of four boundary conditions which determine the four constants B, C, D and F.

We have at $r = a$ the following boundary conditions:

(1) $e_\varphi = 0$;
(2) $e_\varphi^e = 0$;
(3) $e_z = e_z^e$;
(4) $h_\varphi = h_\varphi^e$.

These conditions regard the special features of the helix waveguide: $e_\varphi = 0$ at $r = a$ refers to the assumed perfect conductivity of the helix; h_φ continuous at $r = a$, means that the current in the helix runs strictly in the φ direction. But there is no condition for h_z because the change in h_z on both sides of the helix takes into consideration the connection between the magnetic field and the electric current. But this connection does not determine the relation between the field components on both sides of the helix but gives only the dependence of the excitation constant A of the TE_{01} wave on the wall current. However, this connection is of no interest to the problem we are concerned with. The solution of these four equations is:

$$B = -\frac{1}{2} \frac{1}{\chi_0 R} (1 + a^2\gamma^2)A,$$

$$C = -j \frac{\gamma}{\omega \epsilon_0 \chi_0 R} \left[\frac{\gamma^2}{\chi'^2 a} \frac{W_1(\chi'a)}{W_1'(\chi'a)} + \frac{\epsilon}{\epsilon_0} \frac{\beta_0^2 a}{\chi'} \frac{V_1'(\chi'a)}{V_1(\chi'a)} - \frac{1}{2} \right] A,$$

$$D = -j\omega\mu_0 \frac{\gamma}{\chi_0^2} \frac{a}{R} \frac{J_0(\chi_0 a)}{V_1(\chi'a)} A,$$

$$F = -\frac{\gamma^2}{\chi_0^2} \frac{1}{\chi' R} \frac{J_0(\chi_0 a)}{W_1'(\chi'a)} A.$$

(10)

III. SOME CONCLUSIONS

The equations (4), (7), (8) and (10) give the complete solution to the first order of approximation of the field perturbation of the TE$_{01}$ wave due to curvature of the helix waveguide.

The additional field components (7) and (8) are of the same order of magnitude as a/R and are therefore very small for gentle curvature. This means that the helix waveguide preserves the TE$_{01}$ wave in its general shape even within its curved parts. But the magnitude of the additional field components depends obviously on the construction of the waveguide, the values of ϵ and the distance $(b - a)$ between helix and shield.

If b tends toward a; that is, if the helix waveguide is converted into an ordinary copper pipe with perfectly conducting walls, the TE$_{01}$ wave becomes seriously perturbed. This is shown by the behavior of the constant C. If b tends toward a, $W_1'(\chi'a)$ tends to zero according to (9); this means that C becomes infinitely large. Of course, the perturbation theory has no meaning for infinitely large perturbation terms, but the tendency shows that something is wrong with the TE$_{01}$ wave if the curved helix waveguide tends towards a perfectly conducting ordinary waveguide. The limit of the perturbation method is shown by using the asymptotic approximations for the Hankel functions. This is possible as long as $\epsilon' > 1$ and the TE$_{01}$ wave is far from cutoff:

$$\frac{W_1(\chi'a)}{W_1'(\chi'a)} \approx \cot [\chi'(b - a)],$$

or if $| \chi'(b - a) | \ll 1$,

$$\frac{W_1(\chi'a)}{W_1'(\chi'a)} \approx \frac{1}{\chi'(b - a)}.$$

The perturbation method is therefore applicable as long as

$$\frac{a}{R \, | \chi'(b - a) |} \ll 1.$$

It is interesting to look at the meaning of C in the field pattern. According to (7), C is the excitation constant of the e_z component or, better, of that part of the e_z component which refers to the solution of the homogeneous Maxwell equations and therefore represents a wave of the TM_{11} type. The fact that C tends to infinity with b tending toward a means that the coupling between TE_{01} and TM_{11} waves becomes very strong, as is well known. The coefficient B is not affected by this limit process. That means that, since B is the excitation coefficient of h_z in (7) and therefore represents a TE_{1n} wave, the coupling between TE_{01} waves and TE_{1n} waves is not affected by the conversion of the helix waveguide into an ordinary solid-pipe waveguide. For a real helix waveguide, however, the perturbation terms h and e are small and our first order approximation is valid. It may be expected that there exists an optimum design for the helix waveguide with regard to mode conversion in bend parts, because the constant C is also affected by the choice of ϵ. For instance, for $\epsilon'' \to \infty$, C tends to ∞ too, because of the second term in parenthesis which tends to ∞ like $\sqrt{\epsilon''}$.

It should be mentioned that the additional field components (7) are not simply a superposition of TE_{11} and TM_{11} modes. Though they have the same angular distribution, their radial dependence is different from those of normal modes of the helix waveguide. In order to match these modes to normal modes of the straight guide an infinite series of TE_{1n} and TM_{1n} modes would be needed.

IV. THE ATTENUATION CAUSED BY CURVATURE

According to our assumption of perfectly conducting helix, the TE_{01} wave in the straight waveguide would have no attenuation. But the curvature gives rise to a field in the lossy dielectric between helix and shield. The power dissipation due to this perturbation field causes an attenuation of the whole wave. The power dissipation in the lossy dielectric may be found by calculating the radial power flow through the surface of the helix into that space.

The power dissipation P_d per unit length of the waveguide divided by twice the power P flowing through the cross section of the waveguide in the axial direction gives the attenuation $\alpha_\epsilon = P_d/2P$ (the index ϵ refers to attenuation due to losses in the dielectric). The average power flowing into the dielectric at $r = a$ is:

$$P_d = \frac{1}{2} \operatorname{Re} \int_0^{2\pi} (e_\varphi{}^e \bar{h}_z{}^e - e_z{}^e \bar{h}_\varphi{}^e) a \, d\varphi \qquad \text{at } r = a.$$

The dash over h means the transition to the conjugate complex value. In this way we get

$$\alpha_\epsilon = \frac{1}{2}\frac{|\gamma|}{\chi_0^{2}a}\frac{1}{R^2}\,\text{Re}\left\{\frac{j}{\bar{\chi}'}\left[\frac{|\gamma|^2}{\bar{\chi}'^2}\frac{\bar{W}_1(\chi'a)}{\bar{W}_1'(\chi'a)} - \frac{\bar{\epsilon}}{\epsilon_0}\beta_0^{2}a^2\frac{\bar{V}_1'(\chi'a)}{\bar{V}_1(\chi'a)}\right]\right\}. \quad (11)$$

For TE$_{01}$ far from cutoff and for $\epsilon' > 1$, we have $k'a \gg 1$ and may therefore substitute for the Hankel functions their approximations for large arguments. If we do this we get the much simpler formula:

$$\alpha_\epsilon = \frac{1}{2}\frac{|\gamma|}{a\chi_0^{2}}\frac{1}{R^2}\,\text{Re}\left[\frac{j}{\bar{\chi}'}\left(\frac{|\gamma|^2}{\bar{\chi}'^2} + \frac{\bar{\epsilon}}{\epsilon_0}\alpha^2\beta_0^{2}\right)\cot\bar{\chi}'(b - a)\right]. \quad (12)$$

The transition to the real part gives finally:

$$\alpha_\epsilon = \frac{1}{2}\frac{|\gamma|}{a\chi_0^{2}}\frac{1}{KR^2}$$

$$\frac{\begin{aligned}\cdot\sin u \cos u &\left[\sin\delta\left(\frac{|\gamma|^2}{K^2}\cos 2\delta + \epsilon'\beta_0^{2}a^2\right)\right.\\ &\quad\left.+ \cos\delta\left(\frac{|\gamma|^2}{K^2}\sin 2\delta - \epsilon''a^2\beta_0^{2}\right)\right]\\ + \sinh v \cosh v &\left[\cos\delta\left(\frac{|\gamma|^2}{K^2}\cos 2\delta + \epsilon'a^2\beta_0^{2}\right)\right.\\ &\quad\left.- \sin\delta\left(\frac{|\gamma|^2}{K^2}\sin 2\delta - \epsilon''a^2\beta_0^{2}\right)\right]\end{aligned}}{[(\sin u \cosh v)^2 + (\cos u \sinh v)^2]}, \quad (13)$$

where

$$K^2 = \beta_0^{2}\sqrt{\left[(\epsilon' - 1) + \frac{\chi_0^{2}}{\beta_0^{2}}\right]^2 + (\epsilon'')^2}, \qquad 2\delta = \arctan\frac{\epsilon''}{\epsilon' - 1 + \frac{\chi_0^{2}}{\beta_0^{2}}},$$

$$u = K(b - a)\cos\delta, \qquad v = K(b - a)\sin\delta.$$

Equation (13) gives the attenuation of the TE$_{01}$ wave in the curved helix waveguide due to losses in the dielectric between helix and shield. If ϵ'' tends to zero, δ and v tend to zero and therefore α_ϵ vanishes. In the shielded helix waveguide the additional losses of TE$_{01}$ wave due to curvature may be very small if the losses in the dielectric could be kept small.

Equation (13) may be simplified for two special cases:

i. If b tends to ∞, that is, for the unshielded helix waveguide, the attenuation due to curvature becomes:

$$\alpha_\epsilon = \frac{1}{2}\frac{|\gamma|}{a\chi_0^{2}}\frac{1}{KR^2}\left[\cos\delta\left(\frac{|\gamma|^2}{K^2}\cos 2\delta + \epsilon'a^2\beta_0^{2}\right)\right.$$

$$\left.+ \sin\delta\left(\epsilon''a^2\beta_0^{2} - \frac{|\gamma|^2}{K^2}\sin 2\delta\right)\right]. \quad (14)$$

Now we get an additional attenuation, even for $\epsilon'' = 0$ due to radiation losses into the outer space.

ii. If $\epsilon'' \ll 1$ we get, as long as $0 < K(b - a) < \pi$:

$$\alpha_\epsilon = \frac{1}{4} \frac{a}{R^2} \frac{|\gamma| \beta_0^2}{K\chi_0^2}$$

$$\cdot \frac{K(b - a)\epsilon' - \sin K(b - a) \cos K(b - a)(\epsilon' - 2)}{[\sin K(b - a)]^2} \frac{\epsilon''}{\epsilon' - 1} . \tag{15}$$

Finally, there is some additional curvature attenuation due to the wall currents on the shield and on the helix caused by the additional field components. This attenuation can be neglected when compared with the attenuation caused by losses in all realizable dielectrics.

It may be of interest to mention that the attenuation due to curvature of the helix waveguide is always proportional to $1/R^2$. That is, for the first order of approximation, which considers only terms proportional to $1/R$, there is no attenuation due to curvature. Therefore, the eigenvalue $\chi_0 a$ is unchanged. Instead of calculating the second-order approximation which would also give the attenuation of the TE_{01} wave due to curvature, we have calculated this attenuation by calculating the energy dissipation

V. SOME NUMERICAL EXAMPLES

The following figures show the behavior of the attenuation of TE_{01} mode caused by bends as a function of different parameters:

Fig. 1 gives the attenuation of the TE_{01} mode for fixed frequency and several values of ϵ' as a function of ϵ''. The shield is assumed to be at an
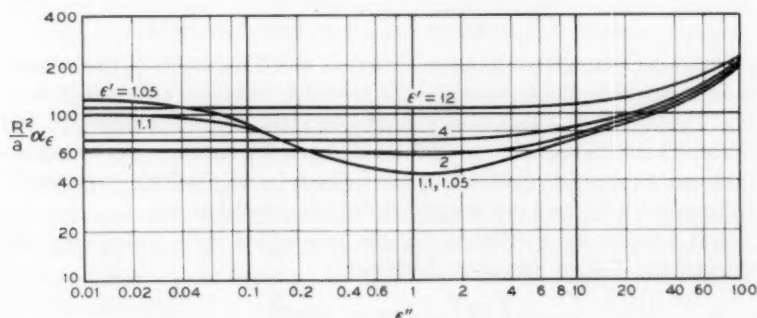
Fig. 1 — Attenuation of the TE_{01} mode due to the curvature of the helix waveguide, with $a\beta_0 = 29.5$.
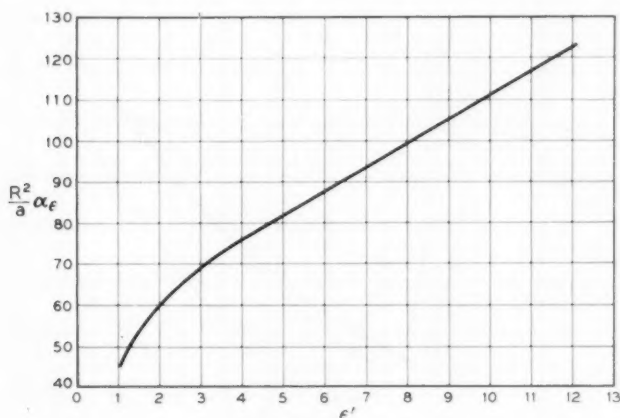
Fig. 2 — Curvature attenuation of the TE$_{01}$ mode, with $\epsilon' = \epsilon''$.

infinite distance $b \rightarrow \infty$. The curves were calculated with the help of (14). The attenuation α is multiplied by R^2/a in order to make it dimensionless. The value of $a\beta_0 = 2\pi(a/\lambda_0) = 29.5$ corresponds to $a = 1$ in. and $f = 55.5$ kmc.

It is interesting that, for greater values of ϵ', the attenuation is nearly constant and depends on ϵ'' only for very large values. Diminishing the value of ϵ' gives a decrease in α. The attenuation has increasing values for decreasing ϵ'' only for values of ϵ' very close to one. The best choice is $\epsilon' = \epsilon''$. It is advisable to make $\epsilon' = \epsilon''$ as close to one as possible, if low values of attenuation are wanted.

Fig. 2 shows the dependence of attenuation on $\epsilon' = \epsilon''$. Let us consider a practical example. If $a = 1$ in., $f = 55.5$ kmc and $\epsilon' = \epsilon'' = 2$:

$$\alpha = \frac{43.3}{R^2} \text{ db/ft}$$

if R is measured in feet. The attenuation of the straight waveguide with copper helix is about $\alpha = 2.92 \times 10^{-4}$ db/ft. The attenuation due to curvature has the same value for a radius of curvature of $R = 385$ ft. $= 117$ meters.

Fig. 3 shows the frequency dependence of curvature attenuation for $\epsilon' = 2$, $\epsilon'' = 2$. If $a = 1$ in., $a\beta_0 = 2\pi(a/\lambda_0) = 18.6$ for $f = 35$ kmc and $a\beta_0 = 40$ for $f = 75$ kmc. In this range, the curvature attenuation at the upper frequency limit is five times the attenuation at the lower frequency limit.
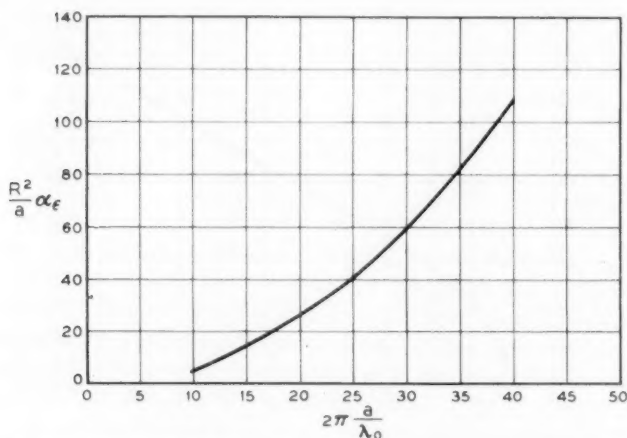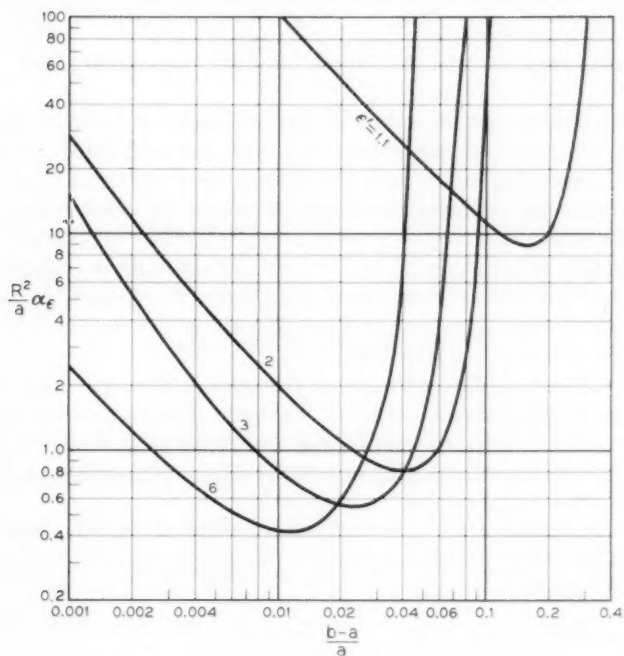
Fig. 3 — Curvature attenuation of the TE_{01} mode, with $\epsilon' = \epsilon'' = 2$.
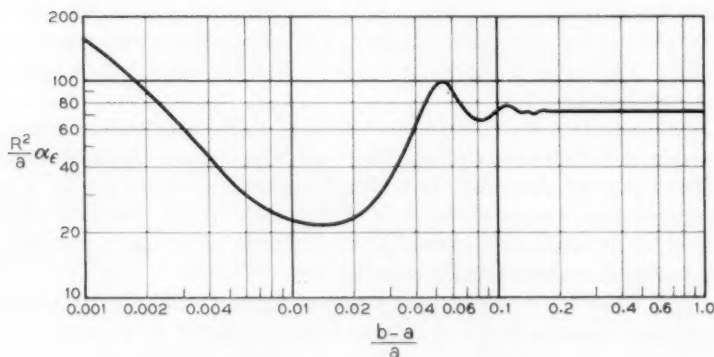
Fig. 4 — Curvature attenuation of the TE_{01} mode, with $\epsilon' = 4$, $\epsilon'' = 2$.

The attenuation due to curvature may very effectively be reduced by proper choice of the distance between helix and shield.

Fig. 4 shows the dependence of attenuation on the relative distance between shield and helix $(b - a)/a$; $\epsilon' = 4$, $\epsilon'' = 2$, $a\beta_0 = 29.5$ in this example. The attenuation first oscillates with increasing $(b - a)/a$ and reaches then a fixed value equal to the value for $b = \infty$ in Fig. 1. But the lowest value is only $\frac{1}{3}$ of the value for $b = \infty$. This favorable effect of the shield can still be improved with decreases in ϵ''. Fig. 5 shows the dependence of $(R^2/a)\alpha$ on $(b - a)/a$ for fixed $\epsilon''/\epsilon' = 0.01$ for different values of ϵ'.

Fig. 5 — Curvature attenuation of the TE_{m} mode, with $\epsilon''/\epsilon' = 0.01$.
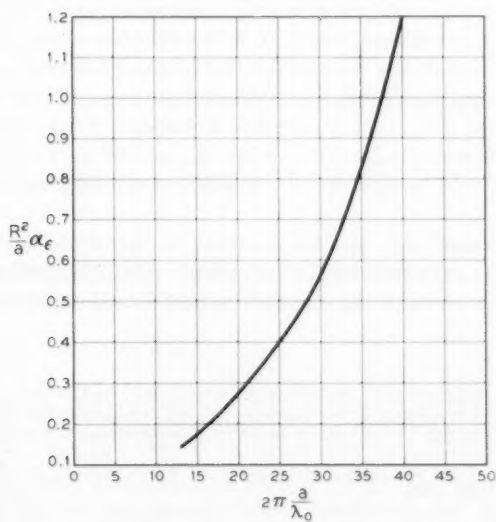
Fig. 6 — Curvature attenuation of the TE_{01} mode, with $\epsilon' = 3$, $\epsilon'' = 0.03$; $(b - a)/a = 0.025$.

It is surprising that the values of the minima are reduced by increase of ϵ'. This is just the opposite behavior as for the unshielded waveguide of Fig. 1. According to (15), α is proportional to ϵ'' for small values. The attenuation may therefore be simply calculated for other values of ϵ'' by means of Fig. 5.

If we again consider $a = 1$ in., $f = 55.5$ kmc and allow α due to curvature to equal the attenuation of TE_{01} mode in straight waveguide, the radius of curvature may be $R = 32.1$ ft. $= 9.8$ meters if $\epsilon' = 6$, $\epsilon'' = 0.06$ and $(b - a)/a = 0.01$.

Proper choice of $(b - a)/a$ allows, therefore, a 10 times larger curvature than in the previous example. That the frequency dependence is not worse for the shielded than for the unshielded guide is shown in Fig. 6. The ratio between the attenuation for $a\beta_0 = 40$ and $a\beta_0 = 18.5$ is again about 5.

If we are only concerned with the attenuation of the TE_{01} mode due to curvature, the shielded waveguide is very much better than the unshielded guide.

VI. SUMMARY

The additional attenuation of a TE_{01} mode due to curvature of the helix waveguide was calculated by means of a perturbation theory. The effects of dielectric constant and of a shield around the helix waveguide were discussed. It was shown that it is advisable to make $\epsilon' = \epsilon''$ as small as possible if low attenuation for a guide without shield and a very thick dielectric coating is wanted. But very much lower attenuation can be obtained if a shield is placed at a proper distance from the helix. To get low attenuation by aid of a shield, the real part of the dielectric constant ϵ' should be as large as possible but its imaginary part as low as possible.

The attenuation due to wall currents in the shield and additional currents due to curvature in the helix may be neglected when it is compared with the losses in the dielectric jacket for all real dielectric materials.

REFERENCES

1. Jouguet, M., Les Effects de la Courbure sur la Propagation des Ondes Dans les Guides a Section Circulatire, Cables et Transmission (Paris), **1**, July 1947, pp. 133–153.
2. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, November 1954, pp. 1209–1265.
3. Rose, C. F. P., Research Models of Helix Waveguides, B.S.T.J., **37**, May 1958, pp. 679–688.
4. Unger, H. G., this issue, pp. 1599–1647.
5. Morgan, S. P. and Young, J. A., Helix Waveguide, B.S.T.J., **35**, November 1956, pp. 1347–1384.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ABRAHAM, R. P. and KIRKPATRICK, R. J.

Transistor Characterization at VHF, Monograph 3007.

ANDERSON, P. W., see Shulman, R. G.

BALLHAUSEN, C. J., see Liehr, A. D.

BARNEY, H. L., see Dunn, H. K.

BASHKOW, T. R.

"Curve-Plotting" Routine for Inverse Laplace Transform of Rational Functions, Monograph 3009.

BENEDICT, T. S. and DURAND, J. L.

Dielectric Properties of BaTiO$_3$ Crystals at Microwave Frequencies, Monograph 3061.

BENN, D. R., see Silverman, S. J.

BOGERT, B. P.

Demonstration of Delay Distortion Correction by Time-Reversal Techniques, Monograph 3006.

BRADY, G. W.

Structure of Sodium Metaphosphate Glass, Monograph 3011.

BRADY, G. W. and KRAUSE, J. T.

Structure in Ionic Solutions, Monograph 3012.

BURRUS, C. A.

Stark Effect from 1.1 to 2.6 Millimeters Wavelength PH$_3$, PD$_3$, DI, and CO, Monograph 3010.

BURRUS, C. A.

Millimicrosecond Pulses in the Millimeter Wave Region, Monograph 3013.

BURRUS, C. A. and GRAYBEAL, J. D.

Stark Effect at 2.0 and 1.2 Millimeters Wavelength: Nitric Oxide, Monograph 3014.

CAGLE, W. B. AND CHEN, W. H.

Designing Low-Level High-Speed Semiconductor Logic Circuits, Monograph 3062.

CHEN, W. H., see Cagle, W. B.

CHYNOWETH, A. G.

Pyromagnetic Effect: A Method for Determining Curie Points, Monograph 3016.

CHYNOWETH, A. G.

Ionization Rates for Electrons and Holes in Silicon, Monograph 3015.

CLOGSTON, A. M.

Inhomogeneous Broadening of Magnetic Resonance Lines, Monograph 3017.

COMPTON, K. G.

Electrical Measurements in Underground Cable Corrosion Problems, Monograph 3063.

COOK, J. S., LOUISELL, W. H. and YOCOM, W. H.

Stability of an Electron Beam on a Slalom Orbit, Monograph 3018.

DEARBORN, E. F., see Nielsen, J. W.

DE COSTE, J. B., HOWARD, J. B. and WALLDER, V. T.

Weathering of Polyvinyl Chloride—Effect of Composition, Monograph 3064.

HERRMANN, G. F.

Optical Theory of Thermal Velocity Effects in Cylindrical Electron Beams, Monograph 3029.

HOWARD, J. B., see De Coste, J. B.

HROSTOWSKI, H. J., see Shulman, R. G.

HUGHES, H. E., WILEY, J. H. and ZUK, P.

Diffused Silicon Diodes—Design, Characteristics, and Aging Data, Monograph 3060.

HUMPHREY, F. B.

Transverse Flux Change in Soft Ferromagnetics, Monograph 3030.

JANSSEN, W. F. and RIGTERINK, M. D.

Microstructure of Ceramics for Communication Equipment, Monograph 3031.

KIRKPATRICK, R. J., see Abraham, R. P.

KISLIUK, P.

Sticking Probabilities of Gases Chemisorbed on the Surfaces of Solids, Monograph 3067.

KRAUSE, J. T., see Brady, G. W.

LEGG, V. E.

Survey of Square-Loop Magnetic Materials, Monograph 3032.

LIEHR, A. D. and BALLHAUSEN, C. J.

Inherent Instability of Octahedral Inorganic Complexes in E_g Electronic States, Monograph 3068.

LOUISELL, W. H., see Cook, J. S.

LOUISELL, W. H. and QUATE, C. F.

Parametric Amplification of Space-Charge Waves, Monograph 3069.

MACKINTOSH, I. M.

Electrical Characteristics of Silicon P-N-P-N Triodes, Monograph 3070.

MANDELL, E. R., see Slichter, W. P.

MC AFEE, K. B., JR.

Stress-Enhanced Diffusion in Glass, Monograph 3033.

MC SKIMIN, H. J.

Elastic Moduli of Single Crystal Germanium as Function of Hydro-static Pressure, Monograph 3034.

MEISSNER, C. R.

High-Vacuum Laboratory for Vapor Deposition of Conductors and Dielectrics, Monograph 2983.

MICHEL, W. S.

Statistical Encoding for Text and Picture Communication, Monograph 3035.

MIRANKER, W. L.

Parametric Theory of $\Delta u + k^2 u = 0$, Monograph 3071.

MONTROLL, E. W. and WARD, J. C.

Quantum Statistics of an Electron Gas, Monograph 3036.

MORRISON, J. A.

Heat Loss of Circular Electric Waves in Helix Waveguides, Monograph 3072.

NIELSEN, J. W. and DEARBORN, E. F.

Growth of Single Crystals of Magnetic Garnets, Monograph 3073.

NOZIERES, P.

Cyclotron Resonance in Graphite, Monograph 3038.

PEARSON, G. L.

Conversion of Solar to Electrical Energy, Monograph 3039.

QUATE, C. F., see Louisell, W. H.

RIORDAN, J.

Combinatorial Significance of a Theorem of Polya, Monograph 3040.

RIGTERINK, N. D., see Janssen, W. F.

ROWE, H. E.

Some General Properties of Nonlinear Elements. II. Small Signal Theory, Monograph 3074.

SHANNON, C. E.

Certain Results in Coding Theory for Noisy Channels, Monograph 2982.

SHULMAN, R. G., WYLUDA, B. J., ANDERSON, P. W. and HROSTOWSKI, H. J.

Nuclear Magnetic Resonance in Semiconductors, Monograph 3075.

SILVERMAN, S. J. and BENN, D. R.

Junction Delineation in Silicon by Gold Chemiplating, Monograph 3041.

SLICHTER, W. P. and MANDELL, E. R.

Molecular Structure and Motion in Irradiated Polyethylene, Monograph 3076.

SOBEL, M., see Gupta, S. S.

SUHL, H.

Origin and Use of Instabilities in Ferromagnetic Resonance, Monograph 3042.

SUHL, H., see Tien, P. K.

TIEN, P. K. and SUHL, H.

A Traveling-Wave Ferromagnetic Amplifier, Monograph 3043.

TUKEY, J. W.

Teaching of Concrete Mathematics, Monograph 3044.

WAITE, T. R.

General Theory of Bimolecular Reaction Rates in Solids and Liquids. Monograph 3045.

WALLDER, V. T., see De Coste, J. B.

WARD, J. C., see Montroll, E. W.

WERNICK, J. H.

Crystal Orientation, Temperature, and Zone Thickness in Zone-Melting, Monograph 2981.

WERTHEIM, G. K.

Transient Recombination of Excess Carriers in Semiconductors, Monograph 3046.

WESTOVER, R. F.

The Thirty Years of Plastics Impact Testing, Monograph 3047.

WILEY, J. H., see Hughes, H. E.

WYLUDA, B. J., see Shulman, R. G.

YEH, SI JUNG and FRISCH, H. L.

Chromatographic Adsorption of Polystyrene, Monograph 3048.

YOCOM, W. H., see Cook, J. S.

ZUK, P., see Hughes, H. E.

Contributors to This Issue

WILLIAM R. BENNETT, B.S. in E.E., 1925, Oregon State College; M.A., 1928, and Ph.D., 1949, Columbia University; Bell Telephone Laboratories, 1925—. His early work was concerned with wire transmission problems and transatlantic submarine cable and coaxial cable telephony. He was later concerned with time-division multiplex telephony and pulse code modulation. For many years he has been engaged in work on noise theory and measurements. He has most recently been concerned with general transmission problems and is technical coordinator in transmission research. Fellow I.R.E.; member A.I.E.E., American Physical Society, Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

O. E. DeLANGE, B.S., 1930, University of Utah; M.A., 1937, Columbia University; Bell Telephone Laboratories, 1930—. Mr. DeLange has concentrated on radio research, first on development of high-frequency transmitters and receivers and frequency modulation research. During World War II he was engaged in radar development. Since then he has been concerned with research in broadband and pulse systems. Member I.R.E.

ALEXANDER FEINER, Technische Hochschule, Vienna; M.S. in E.E., 1952, Columbia University; Bell Telephone Laboratories, 1953—. He has been engaged in design of transistor and scanning circuits, automatic message recording and remote concentration equipment for electronic switching systems. He is in charge of a group concerned with the development of remote line concentrators. Member Sigma Xi.

T. FELDMAN, B.A., 1952, M.A., 1953, and Ph.D., 1955, University of Toronto; Bell Telephone Laboratories, 1956—. He has been concerned with development of networks for electronic switching systems. He was the recipient of the F. J. Orpen Scholarship and the E. F. Burton Fellowship at the University of Toronto. Member Canadian Association of Physicists.

L. FREIMANIS, Technische Hochschule, Munich; B.S. in E.E., 1951 and M.S. in E.E., 1952, Michigan State University; Bell Telephone

Laboratories, 1952—. He first worked on development of signaling systems for Civil Aeronautics Administration networks. Since completing the Communications Development Training Program course he has been engaged in exploratory work on electronic switching systems. Member A.I.E.E., Tau Beta Pi, Eta Kappa Nu, Sigma Pi Sigma.

LEO F. GOELLER, JR., B.E.E., 1953, and M.E.E., 1954, University of Virginia; Bell Telephone Laboratories, 1954—. Since completing the Communications Development Training Program he has worked on scanners and on push-button dialing circuits for electronic switching systems. Member A.I.E.E., I.R.E., Tau Beta Pi.

DIETRICH MARCUSE, Dipl. Vorprufung, 1952, and Dipl. phys., 1954, Berlin Free University; Siemens and Halske (Germany), 1954–1957; Bell Telephone Laboratories, 1957—. At Siemens and Halske, Mr. Marcuse was engaged in transmission research, studying coaxial cable and circular waveguide transmission. His work at Bell Laboratories has been primarily in studies of circular electric waveguides. Member I.R.E.

MIKE PUSTELNYK, B.E., 1950, Youngstown College; M.E., 1953, Rensselaer Polytechnic Institute; Bell Telephone Laboratories, 1953—. As a member of the Radio Research Department, he has been engaged in research and development of PCM systems. Member Sigma Xi.

JOHN W. RIEKE, M.S. in E.E., 1940, Purdue University; Bell Telephone Laboratories, 1940—. Mr. Rieke was engaged in the development of airborne radar for military projects during World War II. He later worked on development of broadband television transmission systems. He has most recently been concerned with development of electronic switching systems. Member A.I.E.E., Tau Beta Pi, Eta Kappa Nu.

HARRISON E. ROWE, B.S., 1948, M.S., 1950, and Sc.D., 1952, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1952—. He was initially associated with a group engaged in systems research. More recently he has been working on mode conversion problems arising in multimode waveguides. Member I.R.E., Sigma Xi, Tau Beta Pi, Eta Kappa Nu.

HOWARD N. SECKLER, B.S. in E.E., 1948, and M.S. in E.E., 1949, Columbia University; Bell Telephone Laboratories, 1949—. He has

specialized in switching research and development and has taken part in work on central office test equipment, switching training and fundamental studies relating to electronic switching. At present he is in charge of a group working on system planning and simulation. Member I.R.E., Tau Beta Pi.

HANS-GEORG UNGER, Dipl. Ing., 1951 and Dr. Ing., 1954, Technische Hochschule, Braunschweig (Germany); Siemens and Halske (Germany), 1951–55; Bell Telephone Laboratories, 1956—. Mr. Unger's work at Bell Laboratories has been in research in waveguides, especially circular electric wave transmission. He holds several foreign patents on waveguides. Member I.R.E., N.T.G. (German Communication Engineering Society).

J. J. YOSTPILLE, S.B., 1948, Massachusetts Institute of Technology; M.E.E., 1955, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1942—. After time out for service in the Navy and studies at M.I.T., Mr. Yostpille rejoined Bell Laboratories in 1948. He first worked on design of toll switching equipment and electronic switching. He later supervised a group engaged in systems planning for electronic switching. He was recently appointed Switching Systems Engineer with responsibility for work on electronic switching. Member I.R.E., Sigma Xi.